Essays on Social Preferences and Choices in Social Dilemmas



Kyeongtae Lee

School of Economics

University of Nottingham

A thesis submitted for the degree of

Doctor of Philosophy

November 2020

Abstract

This thesis presents an empirical examination of decision-making in social dilemmas using online and laboratory experiments.

Chapter 1 provides an overview of the common themes and methodology throughout this research.

Chapter 2 examines the effect of payoff variations on individual decisions using simultaneous prisoner's dilemma games. We focus on three factors that are relevant to payoffs: risk, temptation, and efficiency. We aim to clarify which of these three factors is the most influential determinant of cooperation as there exists mixed evidence so far. We find that temptation has a significantly negative impact on cooperation, whereas risk or efficiency have much weaker effects on cooperation.

Chapter 3 empirically examines the within-subject variability of conditional cooperation across payoff variations using sequential prisoner's dilemma games. We find evidence that conditional cooperation varies with payoffs: subjects are more likely to conditionally cooperate when free-riding causes a higher loss to the first-mover or when free-riding provides a smaller increase in their own earnings. Our empirical findings are in line with theoretical predictions of social preference models, thus it suggests that conditional cooperation reflects underlying social preferences.

Chapter 4 empirically examines the relationship between distributional preferences and conditional cooperation. We estimate individual-level distributional preferences using modified dictator games, then we elicit second-mover strategies using sequential prisoner's dilemma games where payoffs systematically vary. We find that many subjects' strategies, especially conditional cooperation and free-riding, are predicted by their distributional preferences.

Chapter 5 concludes the thesis by summarizing all findings of this research.

Acknowledgements

I would like to express my sincere gratitude to my supervisors Simon Gächter and Martin Sefton for their dedicated guidance and support. Their invaluable advice and encouragement have inspired me to develop my research throughout this thesis.

I would like to extend my appreciation to the members of the Centre for Decision Research and Experimental Economics (CeDEx) for their many useful suggestions for my research. I am indebted to Robin Cubitt, Jose Guinot Saporta, Orestis Kopsacheilis, Chris Starmer, and Fabio Tufano for many helpful comments on my thesis. I also thank my colleagues Malte Baader, Sarah Bowen, Ernesto Gavassa Perez, Anna Hochleitner, Monika Pompeo, Suzanne Robey, Shruti Surachita, and Yuliet Verbel Bustamante for helping me running the experiments at the CeDEx Lab.

Last but not least, I am grateful to my family, Heekyung Song and Jisu Lee, for their patience and encouragement in the course of my PhD.

Contents

1 Introduction **1**
1.1 Social Dilemmas and Social Preferences 1
1.2 Outline of the Thesis 3
References 7

2 Risk, Temptation, and Efficiency in Prisoner's Dilemmas **11**
2.1 Introduction 11
2.2 Related Literature 15
2.2.1 Laboratory Studies 15
2.2.2 Mengel's Meta-study 18
2.2.3 Discussion 19
2.3 Experimental Design and Procedures 20
2.3.1 Experimental Design 20
2.3.2 Experimental Procedures 22
2.4 Results 24
2.4.1 Descriptive Analysis 25
2.4.2 Regression Analysis 28
2.4.2.1 Cooperation and Beliefs 30
2.4.3 Robustness Checks 33
2.5 Discussion and Conclusions 34
References 36
Appendices 40
2.A Robustness Checks 40
2.B Experimental Instructions 46

3 The Variability of Conditional Cooperation in Sequential Prisoner's Dilemmas **50**
3.1 Introduction . . . 50
3.2 Related Literature . . . 55
3.3 Experimental Design & Procedures . . . 57
3.3.1 Experimental Design . . . 57
3.3.2 Experimental Procedures . . . 59
3.4 Results . . . 62
3.4.1 Does Conditional Cooperation Vary with Payoffs? . . . 63
3.4.2 Regression Analysis . . . 65
3.5 Discussion and Conclusions . . . 69
References . . . 71
Appendices . . . 76
3.A Theoretical Analysis . . . 76
3.B Additional Results . . . 81
3.B.1 Second-mover's Decision . . . 81
3.B.2 First-mover's Decision . . . 86
3.C Experimental Instructions . . . 93

4 Distributional Preferences and Conditional Cooperation in Sequential Prisoner's Dilemmas **97**
4.1 Introduction . . . 97
4.2 Related Literature . . . 102
4.3 Estimating Distributional Preferences and Predicting Conditional Cooperation . . . 104
4.3.1 Distributional Preferences . . . 104
4.3.2 Econometric Specification . . . 105
4.3.3 Predicting Conditional Cooperation . . . 106
4.4 Experimental Design and Procedures . . . 109
4.5 Results . . . 115
4.5.1 Estimated Distributional Preferences . . . 115
4.5.2 Second-mover's Strategy . . . 118
4.5.3 Comparing Actual and Predicted Strategies . . . 121
4.5.4 Comparing with Alternative Benchmarks . . . 125
4.5.5 Regression Analysis . . . 127
4.5.6 First-mover's Cooperation . . . 132

4.5.7 Robustness Checks 136
4.6 Discussion and Conclusions 137
References 140
Appendices 145
4.A Distributional Preference Models 145
4.A.1 The Andreoni-Miller Model 145
4.A.2 The Fehr-Schmidt Model 147
4.A.3 The Charness-Rabin Model 150
4.B Robustness Checks 153
4.C Experimental Instructions 226

5 Conclusion **235**
References 239

List of Figures

2.1 The Prisoner's Dilemma (PD) 12
2.2 Aggregate Cooperation Rates with TEMPT, RISK, and EFF . 26
2.3 Comparison of Aggregate Cooperation Rates between PDs . . 27

3.1 The Sequential Prisoner's Dilemma (SPD) 53
3.A1 Predicted Second-mover's Strategies in the CES model 79
3.A2 Predicted Second-mover's Strategies in the FS model 80
3.B1 Aggregate First-mover's Cooperation Rates with TEMPT, RISK, and EFF 87
3.B2 First-mover's Cooperation Rates between PDs 88
3.B3 Average Number of First-mover's Cooperation 90

4.1 The Sequential Prisoner's Dilemma (SPD) 106
4.2 An Example of the Second-mover's Predicted Strategy in the CES model 108
4.3 Three Examples of Subjects' Choices in the Modified Dictator Games 117
4.4 Individual-level Estimates of $\hat{\alpha}$ and $\hat{\rho}$ 118
4.5 Proportions of Conditional Cooperation and Free Riding . . . 120
4.6 Proportions of Second-mover Strategies: Actual and Predicted 121
4.7 Proportions of Second-mover Strategies across Subjects' Types 122
4.8 Comparison between Predicted and Actual Choices 123
4.9 The Hit Rate of the CES model's Prediction at the Individual Level 124
4.A1 Predicted Second-mover's Strategies of the CES model 146
4.A2 Individual-level Estimates of $\hat{\beta}$ 148
4.A3 Predicted Second-mover's Strategies of the FS model 150
4.A4 Individual-level Estimates of $\hat{\lambda}$ and $\hat{\delta}$ 152

List of Tables

2.1 Average Cooperation Rate Regressed on Payoff Indices 19
2.2 Payoff Parameterization for 8 PDs 21
2.3 Determinants of Cooperative Choice in PDs (1) 29
2.4 Determinants of Beliefs in PDs 31
2.5 Determinants of Cooperative Choice in PDs (2) 32
2.A1 Determinants of Cooperative Choice in Prisoner's Dilemmas: K-index, RISK, TEMPT, and EFF 40
2.A2 Determinants of Cooperative Choice in Prisoner's Dilemmas: Fear, Greed, Normalized Loss, and Normalized Gain 43

3.1 Payoff Parameterizations for Sequential Prisoner's Dilemmas . 58
3.2 The Proportion of Conditional Cooperation Strategies 64
3.3 Determinants of Conditional Cooperation 67
3.4 Determinants of Free Riding 68
3.B1 Determinants of Conditional Cooperation 81
3.B2 Determinants of Free Riding 83
3.B3 Determinants of CC (Panel A) and FR (Panel B) 85
3.B4 Determinants of First-mover's Cooperative Choice 92

4.1 Payoff Parameterization for Sequential Prisoner's Dilemmas . 112
4.2 Proportion of Second-mover Strategies 118
4.3 The Proportions of Subjects with a Hit Rate equal or greater of 50% 125
4.4 Determinants of Conditional Cooperation: Symmetric games . 128
4.5 Determinants of Conditional Cooperation: Asymmetric games 130
4.6 Determinants of First-movers' Cooperation: Symmetric games 133
4.7 Determinants of First-movers' Cooperation: Asymmetric games 135
4.B1 Determinants of Conditional Cooperation: Symmetric games 154

4.B2 Determinants of Conditional Cooperation: Asymmetric games 156
4.B3 Determinants of Free Riding: Symmetric games 158
4.B4 Determinants of Free Riding: Asymmetric games 160
4.B5 Determinants of First-mover's Cooperation: Symmetric games 162
4.B6 Determinants of First-mover's Cooperation: Asymmetric games 164
4.B7 Determinants of Conditional Cooperation: Symmetric games & CCEI $>$ 0.80 166
4.B8 Determinants of Conditional Cooperation: Asymmetric games & CCEI $>$ 0.80 168
4.B9 Determinants of Free Riding: Symmetric games & CCEI $>$ 0.80 170
4.B10 Determinants of Free Riding: Asymmetric games & CCEI $>$ 0.80 172
4.B11 Determinants of First-mover's Cooperation: Symmetric games & CCEI $>$ 0.80 174
4.B12 Determinants of First-mover's Cooperation: Asymmetric games & CCEI $>$ 0.80 176
4.B13 Determinants of Conditional Cooperation in the Fehr-Schmidt model: Symmetric games 178
4.B14 Determinants of Conditional Cooperation in the Fehr-Schmidt model: Asymmetric games 180
4.B15 Determinants of Free Riding in the Fehr-Schmidt model: Symmetric games 182
4.B16 Determinants of Free Riding in the Fehr-Schmidt model: Asymmetric games 184
4.B17 Determinants of First-mover's Cooperation in the Fehr-Schmidt model: Symmetric games 186
4.B18 Determinants of First-mover's Cooperation in the Fehr-Schmidt model: Asymmetric games 188
4.B19 Determinants of Conditional Cooperation in the Fehr-Schmidt model: Symmetric games & CCEI $>$ 0.80 190
4.B20 Determinants of Conditional Cooperation in the Fehr-Schmidt model: Asymmetric games & CCEI $>$ 0.80 192
4.B21 Determinants of Free Riding in the Fehr-Schmidt model: Symmetric games & CCEI $>$ 0.80 194

4.B22 Determinants of Free Riding in the Fehr-Schmidt model: Asymmetric games & CCEI > 0.80 196
4.B23 Determinants of First-mover's Cooperation in the Fehr-Schmidt model: Symmetric games & CCEI > 0.80 198
4.B24 Determinants of First-mover's Cooperation in the Fehr-Schmidt model: Asymmetric games & CCEI > 0.80 200
4.B25 Determinants of Conditional Cooperation in the Charness-Rabin model: Symmetric games 202
4.B26 Determinants of Conditional Cooperation in the Charness-Rabin model: Asymmetric games 204
4.B27 Determinants of Free Riding in the Charness-Rabin model: Symmetric games 206
4.B28 Determinants of Free Riding in the Charness-Rabin model: Asymmetric games 208
4.B29 Determinants of First-mover's Cooperation in the Charness-Rabin model: Symmetric games 210
4.B30 Determinants of First-mover's Cooperation in the Charness-Rabin model: Asymmetric games 212
4.B31 Determinants of Conditional Cooperation in the Charness-Rabin model: Symmetric games & CCEI > 0.80 214
4.B32 Determinants of Conditional Cooperation in the Charness-Rabin model: Asymmetric games & CCEI > 0.80 216
4.B33 Determinants of Free Riding in the Charness-Rabin model: Symmetric games & CCEI > 0.80 218
4.B34 Determinants of Free Riding in the Charness-Rabin model: Asymmetric games & CCEI > 0.80 220
4.B35 Determinants of First-mover's Cooperation in the Charness-Rabin model: Symmetric games & CCEI > 0.80 222
4.B36 Determinants of First-mover's Cooperation in the Charness-Rabin model: Asymmetric games & CCEI > 0.80 224

Chapter 1

Introduction

1.1 Social Dilemmas and Social Preferences

This thesis is a contribution to the empirical understanding of behavior in social dilemmas. A social dilemma is a situation in which there exists a conflict between individual and collective interests. Collective action enables everyone to attain higher payoffs, but each individual has a private incentive to unilaterally deviate from the collective action. Many economic phenomena can be viewed as social dilemmas; for example, public goods provision, effort provision in teams, and oligopolistic competition. The simplest form of social dilemma is the prisoner's dilemma, and this has become the focus of much social dilemma research. In the prisoner's dilemma, each agent can either cooperate or defect. Mutual cooperation yields a better outcome than mutual defection, but there exists an incentive to defect because each agent gets a higher payoff by defecting regardless of other's choice.

Experimental prisoner's dilemmas use monetary payoffs to create a game in which subjects maximize their combined earnings by cooperating, but an individual subject maximizes her own earnings by defecting. According to the model that assumes self-interested agents who maximize their material payoffs,

mutual defection is a unique equilibrium. However, experimental evidence clearly shows that some subjects cooperate even in non-repeated interaction (Andreoni, 1995; Clark and Sefton, 2001; Fehr and Gächter, 2000; Fischbacher and Gächter, 2010; Levine, 1998).

Theories of social preferences (Andreoni and Miller, 2002; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), in which an individual's preferences are defined over own and others' earnings, are one of the candidates that can explain cooperation in experimental prisoner's dilemmas. According to social preference models, the material payoffs in a game are transformed into utilities by subjects' social preferences. For subjects who sufficiently care about other's payoffs as well as their own payoffs, defection is no longer a dominant strategy: for example, if subjects are averse to inequalities in material payoffs, the best response to other's choice is conditional cooperation: cooperate if others cooperate, otherwise defect.

Much of the experimental literature has shown the existence of cooperation and social preferences in specific games, but there has been relatively little research investigating the effect of payoff changes across games on cooperation and the linkage between cooperation and social preferences. In this thesis, we examine how cooperation varies with payoffs and whether the variation in cooperation can be explained by models of social preferences. All three chapters, which are joint work with my supervisors Simon Gächter and Martin Sefton, use within-subject designs and employ two-player prisoner's dilemmas where payoffs systematically vary across games. For the payoff variations, our study manipulates the degree of *risk, temptation*, and *efficiency* as defined by Mengel (2018). Risk is the percentage loss from cooperating, rather than defecting, against a defector; temptation is the percentage gain from defecting,

rather than cooperating, against a cooperator; efficiency is the percentage gain from mutual cooperation rather than mutual defection.

1.2 Outline of the Thesis

Chapter 2, entitled *Risk, Temptation, and Efficiency in Prisoner's Dilemmas*, uses an online two-player simultaneous prisoner's dilemma experiment to study the behavioral response to payoff changes across games where measures of risk, temptation and efficiency are systematically manipulated. The main objective of this chapter is to explore which of these three factors is the most influential determinant of cooperation. There exists mixed evidence on which is the most influential in one-shot and stranger-matching games. Ahn et al. (2001) find that greed (a comparable measure to temptation) has a larger effect on cooperation than fear (a comparable measure to risk). Mengel (2018) finds risk best explains the variation in cooperation rates across games, whereas Charness et al. (2016) emphasize the role of efficiency.

To understand the nature of the mixed evidence, we reanalyze data from the existing literature. We find that the measures of risk, temptation, and efficiency used in previous experimental studies are highly correlated, which makes the interpretation of payoff variations in these studies difficult. To resolve this limitation in previous studies, we construct a payoff parameterization in which the measures of risk, temptation and efficiency are orthogonal to each other as payoffs vary across eight prisoner's dilemma games. In the set of games, two distinct levels (*low*, *high*) of each of these payoff indices are used.

We find that increasing temptation reduces cooperation rates significantly. A majority of subjects are sensitive to the level of temptation: those subjects

are more likely to cooperate when the level of temptation is low but more likely to defect when the level of temptation is high. In contrast, increasing risk or efficiency has much weaker effects on cooperation.

Chapter 3, entitled *The Variability of Conditional Cooperation in Sequential Prisoner's Dilemmas*, uses an online two-player sequential prisoner's dilemma experiment to study the variability of *conditional cooperation* across games where payoffs systematically vary. Many previous studies have documented the importance of conditional cooperation, whereby individuals are willing to cooperate if others cooperate, but not willing to cooperate if others defect (Clark and Sefton, 2001; Fischbacher, Gächter and Fehr, 2001; Keser and van Winden, 2000; Kocher et al., 2008).

Several experimental studies have focused on the variability of conditional cooperation over time (Brosig et al., 2007; Muller et al., 2008; Volk et al., 2012) or across different types of games (Eichenseer and Moser, 2020; Mullett et al., 2020), but little research has been conducted to examine the variability of conditional cooperation across payoffs variations of a given game. Theories of social preferences predict that conditional cooperation will vary with payoffs, however, there is little empirical evidence about the extent of this.

We employ the same payoff parameterizations of chapter 2 in which the measures of risk, temptation, and efficiency vary orthogonally across games, but use *sequential prisoner's dilemma* games with the strategy method. In the sequential prisoner's dilemma games, subjects make their cooperate/defect choices in turn, and the second-mover observes the first-mover's choice before making her own. We focus on second-mover's choice across games to measure the variability of conditional cooperation.

We find that conditional cooperation indeed varies with payoffs. The vari-

ation in conditional cooperation is systematically associated with the degree of damage and gain. A subject is more likely to conditionally cooperate when free-riding would impose a higher loss on the first-mover. Also, a subject is less likely to conditionally cooperate when free-riding would provide a higher gain to him/herself. Note that these are within-subject payoff effects, so a given subject who is a conditional cooperator in one payoff parameterization may be a free-rider in another.

Chapter 4, entitled *Distributional Preferences and Conditional Cooperation in Sequential Prisoner's Dilemmas*, examines the relationship between distributional preferences and conditional cooperation in sequential prisoner's dilemmas. A widely used approach to explain unselfish behavior in various contexts uses theories of social preferences. Conditional cooperation also can be explained by social preferences, which incorporate motives such as desires to help others, to equalize earnings, or to maximize combined earnings.

Several studies (Blanco et al., 2011; Teyssier, 2012) empirically examined the relationship between distributional preferences and conditional cooperation, but there are two limitations. First, these studies only used the model of inequality aversion (Fehr and Schmidt, 1999), therefore, these studies cannot examine how efficiency concerns affect conditional cooperation. Second, these studies employed a one-shot social dilemma game with fixed payoff parameters. Thus, the predictive power of distributional preferences for conditional cooperation is not tested under a broad set of payoff parameterizations.

To address these limitations, our study has the following important features. First, we use a CES utility function (Andreoni and Miller, 2002) which can capture fair-mindedness and equality-efficiency tradeoffs. Second, we examine the predictive power of estimated distributional preferences across

twenty-four sequential prisoner's dilemma games with varying payoffs.

We conduct a laboratory experiment consisting of two parts. In Part 1, subjects make an allocation decision in modified dictator games for fifty rounds where the relative price between own and other's earnings vary in each round. The experimental design of Part 1, which is taken from Fisman et al. (2007), allows us to estimate distributional preferences at the individual level. In Part 2, which is conducted one week after Part 1, subjects make decisions in twenty-four sequential prisoner's dilemma games where payoffs vary across games. Part 2 serves as an out-of-sample prediction test to examine whether individual preference estimates can predict second-mover strategies.

The main findings of Chapter 4 are as follows. First, we show that distributional preferences help predict second-mover strategies in the sequential prisoner's dilemmas. Many subjects' second-mover strategies are predicted by their distributional preferences. Second, we find a clear positive relationship between fair-mindedness and conditional cooperation. Subjects who have a higher relative weight on other's earnings are more likely to conditionally cooperate. Third, we also find that fair-mindedness also induces subjects to cooperate as a first-mover. Lastly, we provide strong evidence on the existence and heterogeneity of social preferences across individuals, which is in line with a large body of literature that empirically measure social preferences (Bellemare et al., 2008; Beranek et al., 2015; Blanco et al., 2011; Bruhin et al., 2019; Cabrales et al., 2010; Charness and Rabin, 2002; Dohmen et al., 2009; Fisman et al., 2007; Iriberri and Rey-Biel, 2013).

Chapter 5 concludes by summarizing the results of chapters 2 to 4, discussing directions for future research.

References

Ahn, T. K., E. Ostrom, D. Schmidt, R. Shupp, and J. Walker. "Cooperation in PD games. Fear, Greed, and History of Play." *Public Choice*, 106, 2001, 137–155.

Andreoni, J. "Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments." *Quarterly Journal of Economics*, 110(1), 1995, 1–21.

Andreoni, J., and J. Miller. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, 70(2), 2002, 737–753.

Bellemare, C., S. Kröger and A.V. Soest. "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities." *Econometrica*, 76, 2008, 815–839.

Beranek, B., R. Cubitt, and S. Gächter. "Stated and revealed inequality aversion in three subject pools." *Journal of the Economic Science Association*, 1(1), 2015, 43–58.

Blanco, M., D. Engelmann, and H. T. Normann. "A Within-Subject Analysis of Other-Regarding Preferences." *Games and Economic Behavior*, 72, 2011, 321–338.

Bolton, G.E., and A. Ockenfels. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1), 2000, 166–193.

Brosig, J., T. Riechmann, and J. Weimann. "Selfish in the end? An investigation of consistency and stability of individual behavior." MPRA Paper 2035, University Library of Munich, 2007.

Bruhin, A., E. Fehr, and D. Schunk. "The many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences." *Journal of the European Economic Association*, 17, 2019, 1025–1069.

Cabrales, A., R. Miniaci, M. Piovesan, and G. Ponti. "Social Preferences and Strategic Uncertainty: An Experiment on Markets and Contracts." *American Economic Review*, 100, 2010, 2261–2278.

Charness, G., and M. Rabin. "Understanding Social Preferences with Simple Tests." *The Quarterly Journal of Economics*, 117(3), 2002, 817–869.

Charness, G., L. Rigotti, and A. Rustichini. "Social surplus determines cooperation rates in the one-shot Prisoner's Dilemma." *Games and Economic Behavior*, 100, 2016, 113–124.

Clark, K., and M. Sefton. "The Sequential Prisoner's Dilemma: Evidence on Reciprocation." *The Economic Journal*, 111, 2001, 51–68.

Dohmen,T., A. Falk, D. Huffman, and U. Sunde. "Homo Reciprocans: Survey Evidence on Behavioural Outcomes." *The Economic Journal*, 119, 2009, 592–612.

Eichenseer, M., and J. Moser. "Conditional cooperation: Type stability across games." *Economics Letters*, 188, 2020.

Fehr, E., and K. M. Schmidt. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114, 1999, 817–868.

Fehr, E., and S. Gächter. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90(4), 2000, 980–994.

Fehr, E., and U. Fischbacher. "Social norms and human cooperation." *TRENDS in Cognitive Sciences*, 8(4), 2014, 185–190.

Fischbacher, U., S. Gächter and E. Fehr. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, 71(3), 2001, 397–404.

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

Fisman, R., S. Kariv, and D. Markovits. "Individual Preferences for Giving." *American Economic Review*, 97(5), 2007, 1858–1876.

Iriberri, N., and P. Rey-Biel. "Elicited beliefs and social information in modified dictator games: What do dictators believe other dictators do?" *Quantitative Economics*, 4, 2013, 515–547.

Keser, C., and F. van Winden. "Conditional Cooperation and Voluntary Contributions to Public Goods." *The Scandinavian Journal of Economics*, 102(1), 2000, 23–39.

Kocher, M.G., T. Cherry, S. Kroll, R.J. Netzer, and M. Sutter. "Conditional cooperation on three continents." *Economics Letters*, 101(3), 2008, 175–178.

Levine, D.K. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics*, 1(3), 1998, 593–622.

Mengel, F. "Risk and Temptation: A Meta-Study on Prisoner's Dilemma Games." *The Economic Journal*, 128, 2018, 3182-3209.

Muller, L., M. Sefton, R. Steinberg, and L. Vesterlund. "Strategic behavior and learning in repeated voluntary contribution experiments."*Journal of Economic Behavior and Organization*, 67, 2008, 782–793.

Mullett, T., R. McDonald, and G. Brown. "Cooperation in public goods games predicts behavior in incentive-matched binary dilemmas: Evidence for stable prosociality."*Economic Inquiry*, 58(1), 2020, 67–85.

Teyssier, S. "Inequity and risk aversion in sequential public good games." *Public Choice*, 151, 2012, 91-119.

Volk, S., C. Thöni, and W. Ruigrok. "Temporal stability and psychological foundations of cooperation preferences."*Journal of Economic Behavior and Organization*, 81(2), 2012, 664–676.

Chapter 2

Risk, Temptation, and Efficiency in Prisoner's Dilemmas

2.1 Introduction

A social dilemma is a situation in which there is a conflict between individual and collective interests. Cooperation enables everyone to attain higher payoffs but each individual has a private incentive to unilaterally deviate from cooperation. Many economic phenomena can be viewed as social dilemmas; for example, public goods provision, effort provision in teams, and oligopolistic competition are social dilemmas.

In this study, we focus on the simplest social dilemma, the prisoner's dilemma (henceforth PD), and study how payoffs influence cooperation. In a PD there are two players who simultaneously choose whether to cooperate or defect. The payoff matrix is shown in Figure 2.1 below.

Figure 2.1: The Prisoner's Dilemma (PD)

	Cooperate	Defect
Cooperate	R, R	S, T
Defect	T, S	P, P

Note: $T > R > P > S$ and $2R > T + S$

Rapoport (1967) defines the prisoner's dilemma as satisfying the condition: $T > R > P > S$. Thus, the payoff from mututal cooperation, R, is higher than the payoff from mutual defection, P. However, because $P > S$, cooperation is a "*risky*" choice, exposing the agent to being exploited by a defector. Also, because $T > R$ each agent is "*tempted*" to choose defection as it increases her payoff if the other cooperates. With these parameters, it is a dominant strategy to defect and the unique equilibrium of the game is for both players to defect. We focus on prisoner's dilemma games that satisfy the additional condition $2R > T + S$. This condition ensures that mutual cooperation maximizes combined payoffs.

Experimental prisoner's dilemmas use monetary payoffs to create a game in which subjects maximize their combined earnings by cooperating, but an individual subject maximizes her own earnings by defecting. The model which assumes narrowly self-interested agents, who maximize their own monetary payoffs, predicts both subjects will defect for all payoffs that satisfy $T > R > P > S$. Contrary to this prediction, experimental evidence clearly shows that some subjects cooperate even in non-repeated interaction (e.g., Cooper et al., 1996; Frank et al., 1993; Friedman and Oprea, 2012; Rapoport and Chammah, 1965; Sally, 1995). Rapoport (1967) proposes an index of cooperation, the "K-index" ($\frac{R-P}{T-S}$), to predict how cooperation rates change with

payoffs: the frequency of cooperation is expected to increase with a higher K-index. Subsequent experiment-al studies (e.g., Charness, Rigotti and Rustichini, 2016; Mengel, 2018; Schmidt, Shupp, Walker, Ahn and Ostrom, 2001) show that cooperation rates vary with payoffs, and those empirical findings generally correspond to the prediction of the K-index.

Three factors related to payoffs may affect cooperation in PDs. First, *risk* refers to the danger a cooperator faces of being exploited by a defector: cooperating against a defector leads to a cost of $P - S$. Second, *temptation* refers to the gains to a defector from exploiting a cooperator: defecting against a cooperator leads to a gain of $T - R$. Third, *efficiency* refers to the gain of mutual cooperation: if both choose to cooperate rather than to defect, this leads to a gain of $R - P$ for each subject.

To measure the degree of risk, temptation, and efficiency, we use Mengel's (2018) indices. Mengel (2018) defines RISK ($\frac{P-S}{P}$) as the percentage loss from cooperating against a defector, TEMPT ($\frac{T-R}{T}$) as the percentage gain from defecting against a cooperator, and EFF ($\frac{R-P}{R}$) as the percentage gain from mutual cooperation rather than mutual defection.[1]

This study empirically examines which of these three factors is the most influential determinant of cooperation in PDs. Although the K-index serves as a useful benchmark to explain how cooperation varies with payoffs, the K-index cannot compare the relative effects of risk, temptation, and efficiency on cooperation because the K-index merges these three factors together into a single index. Previous studies investigated the relative effects of distinct motives on cooperation using their own payoff indices and different payoffs, but there is mixed evidence across studies. Ahn et al. (2001) find that "greed"

[1] Other indices have been used in previous studies. See Murnighan and Roth (1983) for a discussion of ten different indices proposed in the previous literature.

($\frac{T-R}{T-S}$, an alternative measure of temptation) has a larger effect on cooperation than "fear" ($\frac{P-S}{T-S}$, an alternative measure of risk). Charness et al. (2016) find that higher payoffs for mutual cooperation are associated with increasing cooperation, whereas Mengel (2018) finds RISK best explains the variation of cooperation rates across one-shot and stranger-matching games where payoffs differ.

One of the possible reasons for contrasting results across studies could be the significant correlations between RISK, TEMPT, and EFF. To resolve this issue, we construct payoff parameterizations so that the indices of RISK, TEMPT, and EFF are orthogonal across games. By introducing two distinct levels *(low, high)* of RISK, TEMPT, and EFF, our experiment consists of eight PDs where these three key variables are uncorrelated. Using this payoff parameterization, we implement a within-subject experiment where all subjects make decisions in eight PDs. At the end of the experiment subjects are paid on the basis of one, randomly selected, game. To rule out confounding factors, such as belief updating, no feedback on any of the individual games is provided until the end of the experiment. Our experimental design allows us to investigate which factor has the largest effect on subjects' decisions.

We find cooperation is significantly lower in the high TEMPT games. Cooperation also increases with EFF, but the effect size is notably smaller than TEMPT. RISK has a very marginal and insignificant effect on cooperation. Overall, our study contributes to the experimental literature that investigates the effect of payoff variations on cooperation and identifies the role of temptation as the most influential factor.

The remainder of this study is organized as follows. In Section 2.2, we review findings from the related literature. Section 2.3 presents our experimental

design and procedures. Section 2.4 presents our results, and we conclude in Section 2.5.

2.2 Related Literature

In this section, we review related studies of the effect of payoff variations on cooperation in prisoner's dilemma experiments. To measure the pure effect of payoff variations on behavior we focus on one-shot games where there is no incentive for cooperation for strategic reasons, rather than on (finitely and infinitely) repeated games.[2] To our knowledge, nine studies examined the effect of payoff variation on cooperation in prisoner's dilemmas. Eight of these implemented laboratory experiments where different payoffs were used in different treatments, with other factors held constant. The last study conducted a meta-analysis based on a large number of prisoner's dilemma studies using different payoff matrices.

2.2.1 Laboratory Studies

Using a between-subjects design, Charness et al. (2016) conducted a one-shot prisoner's dilemma experiment varying R across four treatments. They found that average cooperation rates increase with R. However, note that both EFF and TEMPT change as R changes: the treatment with the greatest R has the highest level of EFF as well as the lowest level of TEMPT, whereas the treatment with the smallest R has the lowest level of EFF and the highest level of TEMPT. Therefore, it is difficult to interpret the result of whether increasing cooperation with increasing R is caused by either increasing effi-

[2] For a discussion of experiments with infinitely repeated games, see Dal Bó and Fréchette (2018).

ciency or decreasing temptation or both. Our design avoids this ambiguity using orthogonal levels of RISK, TEMPT, and EFF.

The following seven studies implemented within-subject experiments that consist of multiple prisoner's dilemmas with varying payoffs. Engel and Zhurakhovska (2016) studied 11 one-shot PDs where P varied across games and T, S and R were held constant. Each subject played all 11 PDs with no feedback between games. The authors found that cooperation decreases as P increases. Note, however, that TEMPT does not vary across the games, and the observed decrease in cooperation may be due to either increasing RISK or decreasing EFF.

Three studies used designs in which subjects played a series of games against randomly changing opponents, with payoffs varying across games. In Vlaev and Chater (2006) the K-index varied across rounds and the cooperation rate increased with the K-index. Ahn et al. (2001) and Schmidt et al. (2001) examined the impact of variations in greed, $(\frac{T-R}{T-S})$, and fear, $(\frac{P-S}{T-S})$, on cooperation. These two studies are closely related to our own as greed and fear are alternative measures of temptation and risk (based on a different normalisation to those used in the TEMPT and RISK indices).[3] Ahn et al. (2001) varied the payoffs across four games by using *high* and *low* values of T and S but holding R and P constant, whereas Schmidt et al. (2001) varied the values of R and P across six games while keeping the values of T and S constant. These studies both found a significant effect of greed, but whereas Ahn et al. (2001) found that greed has a greater impact than fear on cooperation, Schmidt et al. (2001) reported similar effect sizes of greed and fear on cooperation. Note that all these three studies provided feedback during the

[3] Note that greed and fear do not completely disentangle the roles of temptation and risk since both indices are influenced by variations in T and S through their denominators.

experiment, and therefore cooperation might be affected by the outcome of previous games as well as by payoff changes. (Indeed, the effect of feedback across game is the focus of Vlaev and Chater, and they report significant feedback effects.) Our design does not provide any feedback until the end of the experiment to exclude the effect of feedback on decision-making. Our study also differs from these studies in that we use Mengel's (2018) indices of RISK, TEMPT, and EFF.

Au et al. (2012) and Ng and Au (2016) define the relative risk of cooperation (henceforth riskiness) as $\frac{(R-S)}{(R-S)+(T-P)}$, and examine how riskiness and subjects' risk attitudes affect cooperation. Au et al. (2012) employed 18, 16, and 28 PDs in three experiments, while Ng and Au (2016) used 24 PDs. No feedback was provided until the end of the experiment in either study. Both studies found that the effect of riskiness of PDs depends on subjects's risk attitude: risk-averse subjects are more likely to cooperate in a less risky game, while risk-seeking subjects are more likely to cooperate in a riskier game. However, the measure of riskiness does not disentangle risk, temptation, and efficiency: riskiness increases as T decreases or R increases. Therefore, increasing cooperation of risk-seeking subjects with increasing riskiness might be caused by either decreasing temptation or increasing efficiency or both.

Lastly, a recent experimental study (Weber, mimeo) employed 17 one-shot PDs, varying RISK, TEMPT, and EFF with no feedback on outcomes until the end of the experiment. The focus of this study was the relationship between individual cooperation and social value orientation under varying payoff conditions. Weber found that after controlling for social value orientation, increasing TEMPT significantly reduces cooperation, and increasing EFF significantly increases cooperation. This study also found that RISK does not

have a significant effect on cooperation. Our design differs from this study in that, in our experiment, variations in RISK, TEMPT, and EFF are strictly orthogonal across treatments.

Overall, it is inconclusive from the findings of previous laboratory experiments which factor out of risk, temptation, and efficiency is the most influential determinant of cooperation.

2.2.2 Mengel's Meta-study

A recent meta-study (Mengel, 2018) used extensive data from existing prisoner's dilemma studies to clarify how much of the variation in average cooperation rates across studies can be explained by RISK, TEMPT, and EFF.

We reanalyze the data of Mengel because some of the games do not meet our PD criteria ($T > R > P > S$, and $2R > T + S$). We report the results, using the same reduced-form specification as in Mengel (2018), in Table 2.1. Columns (1)-(4) show the results of OLS regressions in which the dependent variables are the average cooperation rates in one-shot and stranger matching games. Column (1) consists of all one-shot and stranger-matching games in Mengel's data set. Mengel also reported the results separately for studies with non-negative payoffs: these are reported in column (2). In column (3), we exclude the games which do not satisfy the condition of ($T > R > P > S$, and $2R > T + S$) from column (1). In column (4), we additionally dropped two games which have negative payoffs.

In columns (1)-(2), RISK and EFF are significantly associated with cooperation in social dilemmas, which reproduces the results of Mengel (2018, Table 1, p. 3193). However, these findings change in columns (3)-(4). As we restrict the sample to those studies that satisfy our PD conditions, the effect

of RISK on cooperation becomes insignificant. The index of efficiency (EFF) is significant in all specifications, but the coefficient varies substantially across regressions. Notably, TEMPT is not significant in any of the regressions.

Table 2.1: Average Cooperation Rate Regressed on Payoff Indices

	(1) All data	(2) All data satisfying $S \geq 0$	(3) All data satisfying $T > R > P > S$ & $2R > T + S$	(4) Same as (3) & $S \geq 0$
RISK	-0.255***	-0.266***	-0.045	-0.127
	(0.061)	(0.058)	(0.123)	(0.125)
TEMPT	0.003	-0.013	-0.492	-0.283
	(0.080)	(0.074)	(0.305)	(0.300)
EFF	0.291***	0.388***	0.301*	0.469***
	(0.089)	(0.087)	(0.149)	(0.159)
Constant	0.370***	0.343***	0.304**	0.236*
	(0.084)	(0.080)	(0.130)	(0.136)
Adj R^2	0.350	0.451	0.167	0.295
Obs.	73	71	36	34

Notes. Data source: Mengel (2018; stranger and one-shot games). All columns are OLS regressions, and standard errors in parentheses. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

2.2.3 Discussion

Two key findings emerge from this summary of related studies. First, correlations between the key variables of RISK, TEMPT, and EFF make the interpretation of treatment effects in these studies difficult. For example, there is a strong negative correlation between TEMPT and EFF in the games used in Charness et al. (2016), a strong negative correlation between RISK and EFF in Schmidt et al. (2001) and Au et al. (2012), and a strong negative correlation between RISK and TEMPT in Ng and Au (2016). Second, the results from combining data from multiple studies are sensitive to whether non-negative

payoffs are used and whether mutual cooperation leads to maximal combined earnings.

Based on these observations, we design an experiment that has the following important features. First, we construct payoff parameterizations in which the variables of interest – RISK, TEMPT, and EFF indices – are orthogonal. Second, we restrict attention to PDs with non-negative payoffs where mutual cooperation maximizes combined earnings. In addition, we use a within-subject design where each subject plays multiple PDs, with no feedback until the end of the experiment, to examine how individual behavior changes with payoffs. The next section provides details of the experimental design and procedures.

2.3 Experimental Design and Procedures

2.3.1 Experimental Design

The experiment is based on a simple two-player prisoner's dilemma game. Our main interest lies in exploring whether RISK, TEMPT, or EFF is the most influential determinant of cooperation. To construct orthogonal payoff indices in the simplest way, we set up two distinct levels – *low* and *high* – of RISK, TEMPT, and EFF, respectively.

Table 2.2 shows the payoff parameterization used in the experiment. The experiment consists of eight PDs that satisfy the conditions i) $T > R > P > S$, ii) $2R > T + S$, and iii) $S > 0$. With these constraints, $0 <$ RISK < 1, $0 <$ EFF < 1, and $0 <$ TEMPT < 0.5.[4] We chose payoffs so that the *low* and *high* levels of RISK were 0.55 and 0.90, the *low* and *high* levels of TEMPT were

[4] The upper bound of 0.5 for TEMPT is implied by the restriction $2R > T + S$.

0.17 and 0.38, and the *low* and *high* levels of EFF were 0.20 and 0.60. Thus, the high levels of TEMPT and RISK are approximately double the low levels, while for EFF the high level is triple the low level. The K-index is determined by the level of RISK, TEMPT, and EFF, and varies across games. Game 1 (*low* RISK, *low* TEMPT, *high* EFF) has the highest K-index of 0.59, while Game 8 (*high* RISK, *high* TEMPT, *low* EFF) has the lowest K-index of 0.13.

Table 2.2: Payoff Parameterization for 8 PDs

Game	T	R	P	S	RISK	TEMPT	EFF	K-index
G1	600	500	200	90	0.55	0.17	0.60	0.59
G2	600	500	200	20	0.90	0.17	0.60	0.52
G3	800	500	200	90	0.55	0.38	0.60	0.42
G4	800	500	200	20	0.90	0.38	0.60	0.38
G5	600	500	400	180	0.55	0.17	0.20	0.24
G6	600	500	400	40	0.90	0.17	0.20	0.18
G7	800	500	400	180	0.55	0.38	0.20	0.16
G8	800	500	400	40	0.90	0.38	0.20	0.13

Note: RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: efficiency $\frac{R-P}{R}$, K-index: $\frac{R-P}{T-S}$.

The payoffs were also chosen to be strictly positive multiples of ten in order to avoid zero or non-rounded payoffs. R (500) is constant across all PDs, while our experiment has two distinct values of T (600, 800) and P (200, 400), and four distinct values of S (20, 90, 40, 180).

The implications of payoff indices' values are as follows. The value 0.55 (0.90) of RISK implies that an individual who cooperates against a defector will incur a 55% (90%) loss of monetary payoff compared to the payoff from mutual defection. The value 0.17 (0.38) of TEMPT implies that an individual who defects against a cooperator will accrue a 17% (38%) gain in monetary payoff

compared to the payoff from mutual cooperation. The value 0.60 (0.20) of EFF implies that an individual will accrue a 60% (20%) gain in monetary payoff if both subjects cooperate compared to the payoff from mutual defection.

2.3.2 Experimental Procedures

We conducted the online interactive experiment in Spring 2019 using MTurk. Subjects were residents of the United States. We conducted five sessions with a total of 160 participants. None of the subjects participated in more than one session. After reading the instructions and passing the control questions, each participant was paired with another subject.

Each pair played all eight games of Table 2.2. To control for potential order effects, we randomized the sequence of games at the pair level. Subjects completed two tasks in each game: a decision task and a belief elicitation task. Both the decision task and the belief elicitation task were presented on the same screen. In the *decision task*, they simply chose whether to cooperate or to defect. Decisions were neutrally labelled as options A and B, with labeling randomly chosen at the pair level in each game so that the cooperative decision was labeled A in some games and B in others. In the (non-incentivized) *belief elicitation task* we asked subjects to report the probability (between 0 and 100 percent) that their opponent will choose option A. Subjects could complete the tasks in either order, but to control for potential presentation effects, we randomized whether the decision task or belief task appeared at the top of the screen at the pair level.

For each PD, subjects had to answer eight additional control questions about the payoffs before making decisions or reporting beliefs. These additional control questions were intended to ensure that subjects understood the

implications of their decisions and recognized the payoff changes across games.

Subjects did not receive any feedback on the other's choice or the outcome of each PD until the end of the experiment. Once subjects completed the tasks for all games, we asked them to complete a short post-experimental questionnaire eliciting basic demographic information. This included a self-assessment of subjects' political orientation, using an 11-point left-right scale that is frequently used in surveys (e.g., Costa Lobo and Curtice, 2015; Kroh, 2007).

We implemented the experiment using the software LIONESS (Giamattei, Yahosseini, Gächter and Molleman, 2020). Subjects were paired with another participant on a real-time basis and they conducted each task at the same time. This implies a subject had to wait until the opponent made a decision to proceed to the next game. As subjects needed to wait until their opponent made a decision, long waiting times could increase the risk of reduced attention. We took the following measures to retain attention and encourage successful completion of the experiment. Before participants entered the experiment, we told them to avoid distractions during the experiment. In addition, participants who were inactive for more than 30 seconds (i.e. no mouse movement or no keyboard input) got an alert voice message and a blinking text on their browser. If an inactive participant did not respond to the alert message for a further 30 seconds, such an inactive participant was removed from the experiment and the remaining person was able to continue the experiment.

At the end of the experiment, one out of eight games was randomly chosen for payment. Subjects were reminded of their choices and informed about the outcome for the randomly chosen game. If both subjects completed the entire experiment, they were paid according to the outcome of the randomly chosen

game. If one of the pair had dropped out during the experiment, the computer randomly selected the payoff-relevant game for the remaining subject. Then the computer randomly selected one out of four monetary outcomes (i.e. T, R, P, or S) of the chosen game for payment to the remaining subject. We explained this payment scheme clearly in the instructions.

As is commonplace in online experiments, there was a non-negligible attrition rate: 39 out of 160 subjects (24%) dropped out during the experiment.[5] For subjects who completed the experiment, the average age was 34 years (between 18 and 64 years) and 50% were female. Subjects' earnings ranged from \$1.20 to \$9.00, averaging \$5.00. On average, the experiment lasted about 30 minutes, including the completion of a post-experimental questionnaire. Subjects were informed of their payment immediately upon completion of the experiment and were paid within 24 hours.

2.4 Results

The following analysis is structured to discuss our main research questions: First, are subjects sensitive to payoff variations in their decision-making? If so, which variable among risk, temptation, and efficiency affects cooperation most? Lastly, are our results robust to using other indices? For our analysis, we only include the decisions of subjects who completed the experiment: thus our data set consists of 968 observations (121 subjects $\times$ eight games).

[5] The dropout rate in our experiment is not too different from that of similar interactive online experiments. For example, Arechar et al. (2018) report a 20% dropout rate in their interactive 4-player public goods game.

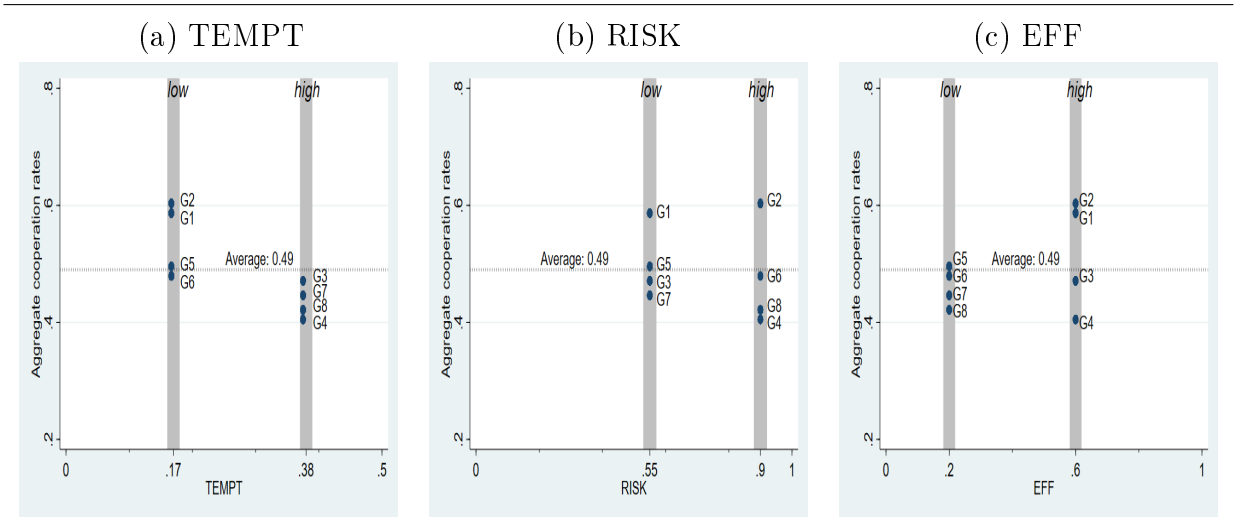
2.4.1 Descriptive Analysis

We examine how many subjects change their decision across eight games: 17% always defect, 13% always cooperate, and the remaining 70% of subjects choose cooperation and defection at least once. We refer to the 70% of subjects who change their decisions across games as "switchers" for the rest of the analysis. On average, switchers cooperate in four out of eight games. The finding that a majority of subjects change their decisions across games suggests that variation in RISK, TEMPT, or EFF might be significantly associated with individual cooperation.

Next, we review aggregate cooperation rates across eight games. The average cooperation rate is 49%. Games 1 and 2 (which both have low TEMPT and high EFF) show significantly higher cooperation rates than the other games, while cooperation rates across G3-G8 are not significantly different.[6] Figure 2.2 illustrates aggregate cooperation rates depending on the level of each payoff index. Aggregate cooperation rates in the *low* TEMPT is always greater than in the *high* TEMPT (Figure 2.2(a)). For RISK (Figure 2.2(b)) and EFF (Figure 2.2(c)), the relationship between aggregate cooperation rates and these indices is less clear compared to TEMPT: the variance of cooperation rates increases as RISK or EFF increases.

[6] We can strongly reject the hypothesis that all games have a similar effect on cooperation (Cochran's Q test, $p = 0.0001$). However, we cannot reject this hypothesis if we exclude games 1 and 2.

Figure 2.2: Aggregate Cooperation Rates with TEMPT, RISK, and EFF

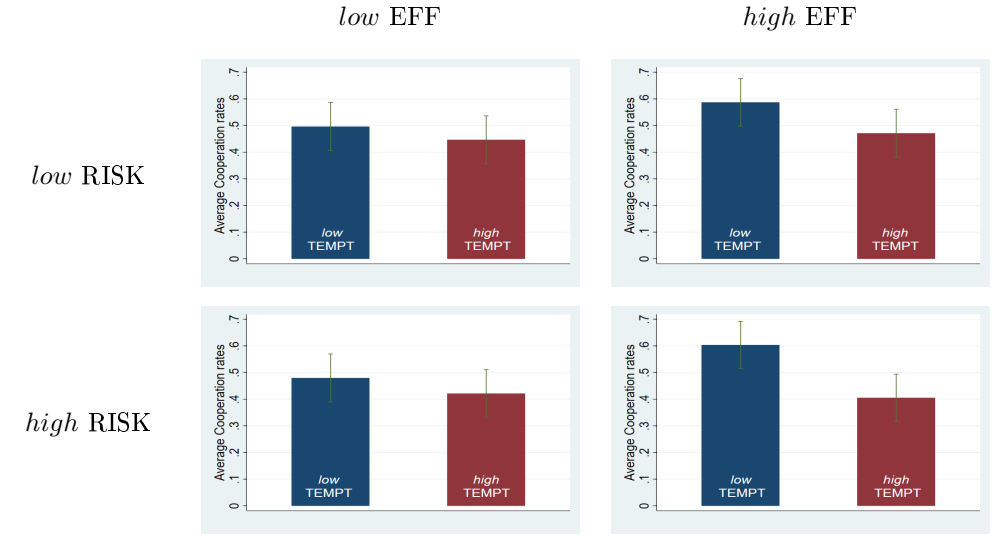


Note: TEMPT: $\frac{T-R}{T}$, RISK: $\frac{P-S}{P}$, EFF: $\frac{R-P}{R}$.

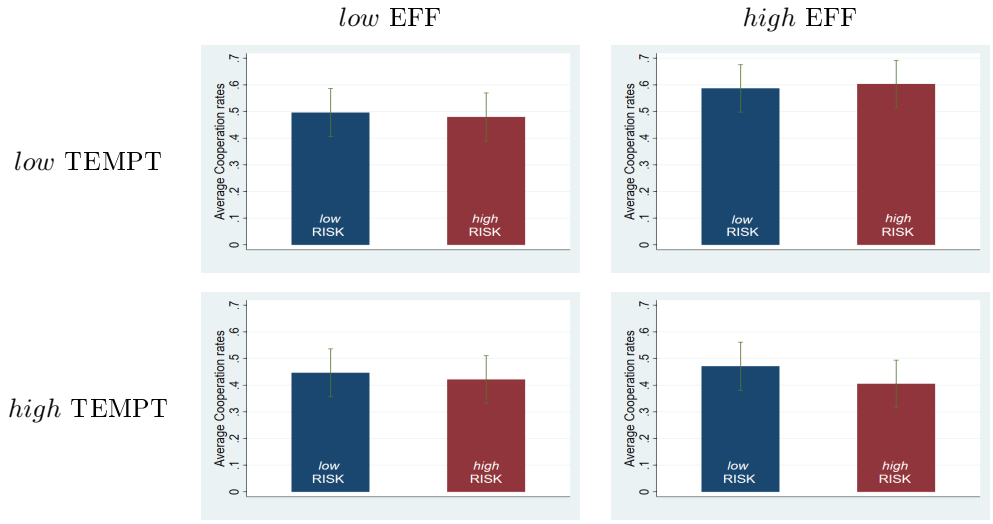
To examine how cooperation rates are affected by the variation in each payoff index holding other payoff indices constant, we compare the games one by one (Figure 2.3). First, we find that average cooperation rates vary more with TEMPT (Figure 2.3(a)) than with RISK (Figure 2.3(b)) or EFF (Figure 2.3(c)). In Panel (a), we see that cooperation rates for the *low* TEMPT are on average 11 percentage points higher than for the *high* TEMPT (with differences ranging from 5 percentage points to 20 percentage points depending on the levels of the other indices). In contrast, cooperation rates for the *low* RISK are only on average 3 percentage points higher than for the *high* RISK (with differences ranging from 2 percentage points to 7 percentage points). Cooperation rates for the *high* EFF are 6 percentage points higher than for *low* EFF (with differences ranging from 2 percentage points to 12 percentage points).

Figure 2.3: Comparison of Aggregate Cooperation Rates between PDs

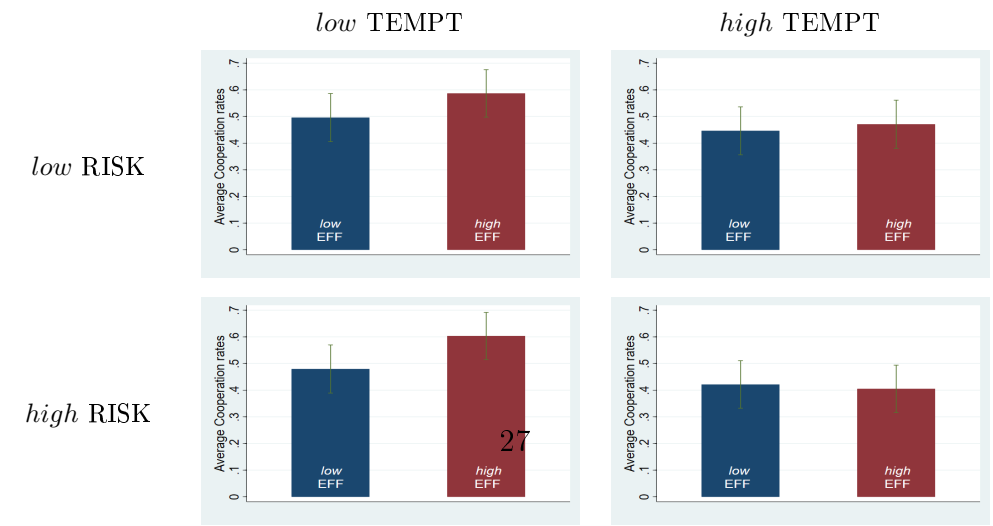
Panel (a). *low* vs *high* TEMPT



Panel (b). *low* vs *high* RISK



Panel (c). *low* vs *high* EFF



Second, the effect of TEMPT is substantially affected by the level of EFF. Increasing TEMPT leads to a greater reduction in cooperation rates when EFF is *high* than when EFF is *low*. Third, the effect of EFF is also substantially affected by the level of TEMPT. Increasing EFF leads to a greater increase in cooperation rates when TEMPT is *low* than when TEMPT is *high*. Lastly, increasing RISK has marginal effects on cooperation rates regardless of the levels of TEMPT and EFF.

2.4.2 Regression Analysis

In this section, we use a regression approach to analyze the relationship between individual decisions to cooperate and payoff indices. Table 2.3 shows the results of random effect probit regressions: Column (1) examines how cooperation varies with the K-index, while Column (2) reports the effects of the RISK, TEMPT, and EFF payoff indices. Columns (3) and (4) provide corresponding regression results for the subsample of switchers.

First of all, the K-index has a highly significant effect on cooperation. Based on the complete sample, the marginal effect of K-index is 0.296, which implies that the probability of cooperation increases by about 3 percentage points when the K-index rises by 0.1 points. The finding that the cooperation rate increases with the K-index is in line with the existing PD literature (e.g., Balliet and Van Lange, 2013; Moisan et al., 2018; Murphy and Ackermann, 2015, Vlaev and Chater, 2006).

Second, the signs of coefficients for RISK, TEMPT, and EFF correspond to the prediction of the K-index: the coefficients of RISK and TEMPT are negative and the coefficient of EFF is positive. Without considering their significance, these coefficients imply that subjects are more likely to cooperate

when (1) the degree of risk decreases, (2) the degree of temptation decreases, or (3) the degree of efficiency increases.

Table 2.3: Determinants of Cooperative Choice in PDs (1)

	(1) All	(2) All	(3) Switcher	(4) Switcher
K-index	0.296***		0.434***	
	(0.080)		(0.112)	
RISK		-0.065		-0.098
		(0.061)		(0.088)
TEMPT		-0.530***		-0.765***
		(0.114)		(0.154)
EFF		0.138**		0.206**
		(0.069)		(0.099)
Round	-0.014**	-0.015***	-0.020**	-0.022***
	(0.006)	(0.006)	(0.009)	(0.008)
Observations	968	968	680	680
Log-likelihood	-536.5	-531.1	-446.0	-440.6
BIC	1,100.5	1,103.5	918.2	920.4
Pseudo R^2	0.019	0.029	0.024	0.036

Notes: Average marginal effects from random effect probit regression with standard errors clustered on individuals. K-index: $\frac{R-P}{T-S}$, TEMPT: $\frac{T-R}{T}$, RISK: $\frac{P-S}{P}$, EFF: $\frac{R-P}{R}$. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Third, switchers are more sensitive to payoff variations (columns (3) and (4)). For switchers, who represent 70% of subjects, the effect sizes of the K-index (0.434), and the RISK (-0.098), TEMPT (-0.765), and EFF (0.206) indices are approximately 1.5 times greater than the effect sizes based on the complete sample.

Fourth, TEMPT has a significantly larger impact on cooperation than RISK and EFF. In the regression based on the complete sample (Column (2)) the marginal effect of TEMPT is -0.530, compared with much smaller effect sizes for EFF (0.138) and RISK (-0.065), and TEMPT and EFF are significant at the 1% and 5% levels, respectively, whereas RISK is not significant. Similarly, for switchers (Column (4)) TEMPT has a much larger effect size than

EFF or RISK, and again TEMPT and EFF are significant at the 1% and 5% levels respectively, while RISK is insignifcant.

Lastly, we observe a mild decreasing round effect: the probability of cooperation decreases by 1.4 - 2.2 percentage points from one game to the next.

2.4.2.1 Cooperation and Beliefs

We consider two channels by which the indices may affect cooperation. First, a player's belief about the opponent's choice could be influenced by payoffs, and this in turn may affect the player's own choice. Second, even if a player's beliefs do not change in response to a change in payoff, the change in payoff may directly affect a player's willingness to cooperate. In this subsection, we examine these two channels.

In Table 2.4, we examine how payoff indices affect a player's subjective beliefs about the probability with which the opponent will cooperate. Columns (1)-(2) provide the results for all subjects, and columns (3)-(4) provide the results for the subsample of switchers. Column (1) shows that beliefs that the opponent will cooperate increase with the K-index: beliefs increase by about 1.6 percentage points when the K-index increases by 0.1 point, and this effect is significant at the 1% level. Column (2) shows that this is driven by the significant effects of TEMPT and EFF. A 0.1 point decrease in TEMPT increases belief by about 2 percentage points, while a 0.1 point increase in EFF increases belief by about 1 percentage point. Both these effects are significant at the 5% level. RISK has an insignificant effect on subjective beliefs.

Columns (3) and (4) shows that the effect sizes are larger for the subsample of switchers, but the overall pattern is similar. Beliefs are more sensitive to TEMPT than to EFF or RISK, and TEMPT is now significant at the 1% level,

whereas EFF is still significant at the 5% level and RISK is still insignificant.

In all regressions we find a decreasing round effect: beliefs about opponent's cooperation decay 1.2-1.5 percentage points from one game to the next, and this effect is significant at the 1% level.

Table 2.4: Determinants of Beliefs in PDs

	(1) All	(2) All	(3) Switcher	(4) Switcher
K-index	0.157***		0.188***	
	(0.048)		(0.061)	
RISK		0.002		0.042
		(0.053)		(0.064)
TEMPT		-0.201**		-0.275***
		(0.080)		(0.103)
EFF		0.099**		0.119**
		(0.042)		(0.053)
Round	-0.012***	-0.013***	-0.014***	-0.015***
	(0.004)	(0.004)	(0.005)	(0.005)
Observations	968	968	680	680
Log-likelihood	-124.9	-123.6	-95.0	-92.7
BIC	284.2	295.3	222.7	231.0
R^2	0.016	0.018	0.023	0.029

Notes: Random effects panel regression with standard errors clustered on individuals. K-index: $\frac{R-P}{T-S}$, TEMPT: $\frac{T-R}{T}$, RISK: $\frac{P-S}{P}$, EFF: $\frac{R-P}{R}$. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Next, we examine how payoff indices and a player's subjective beliefs affect cooperation. Table 2.5 reports the results of random effects probit regressions of the decision to cooperate on payoff indices, controlling for beliefs. Again we report separate regressions based on the complete sample (Columns (1)-(2)) and the subsample of switchers (Columns (3)-(4)).

First of all, subjective beliefs about the opponent's cooperativeness are significantly associated with cooperation. As beliefs increase, the probability of cooperation increases: the coefficients range from 0.377 to 0.424 across

columns (1) - (4), which implies that the probability of cooperation increase by 3.8 $\sim$ 4.2 percentage points as beliefs increase by 10 percentage points. The significant relation between belief and choice is in line with the findings of previous social dilemma studies (e.g., Croson, 2007; Gächter and Renner, 2018; Neugebauer, Perote, Schmidt and Loos, 2009; Gächter, Nosenzo, Renner and Sefton, 2012).

Second, even after controlling for beliefs, the K-index and TEMPT have a significant effect on cooperation. Similar to the regression results without controlling beliefs, the probability of cooperation increases with the K-index but decreases with TEMPT. After controlling for beliefs, cooperation is not significantly affected by RISK or EFF.

Table 2.5: Determinants of Cooperative Choice in PDs (2)

	(1) All	(2) All	(3) Switcher	(4) Switcher
K-index	0.254***		0.357***	
	(0.085)		(0.112)	
RISK		-0.084		-0.121
		(0.065)		(0.086)
TEMPT		-0.490***		-0.669***
		(0.115)		(0.147)
EFF		0.105		0.154
		(0.074)		(0.099)
Round	-0.010	-0.012*	-0.014*	-0.017**
	(0.006)	(0.006)	(0.009)	(0.008)
Belief	0.385***	0.377***	0.424***	0.415***
	(0.072)	(0.072)	(0.081)	(0.081)
Observations	968	968	680	680
Log-likelihood	-508.6	-504.2	-425.1	-420.5
BIC	1,051.7	1,056.5	882.7	886.5
Pseudo R^2	0.070	0.078	0.070	0.080

Notes: Average marginal effects from random effect probit regression with standard errors clustered on individuals. K-index: $\frac{R-P}{T-S}$, TEMPT: $\frac{T-R}{T}$, RISK: $\frac{P-S}{P}$, EFF: $\frac{R-P}{R}$. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

In summary, we find that TEMPT is the most influential determinant of cooperation and it has both *direct* and *indirect* effects on cooperation. Increasing TEMPT directly deters cooperation in the sense that even if a subject's expectations about the opponent's choice does not change, subjects are less willing to cooperate. In addition, increasing TEMPT has an indirect effect discouraging cooperation because it reduces a subject's expectation of their opponent's cooperativeness, and this makes them less likely to cooperate. Increasing EFF also increases cooperation, but the effect size is much smaller and works only through this indirect channel that increasing EFF strengthens the belief about the opponent's cooperativeness.

2.4.3 Robustness Checks

So far we have examined the effect of risk, temptation, and efficiency by using Mengel's (2018) payoff indices. In Appendix A, we report a number of robustness checks. First, we show that these results are robust to the inclusion of session dummies and additional control variables (Table A1). We find that the effect sizes and significance of payoff indices are essentially unchanged, and that session dummies (Wald test, $\chi^2 = 4.10$, $p = 0.392$), individual characteristics (i.e. age, gender, political orientation, MTurk experience, employment, education, income level, ethnicity) ($\chi^2 = 8.08$, $p = 0.621$), and task characteristics (labeling of cooperative choice as A or B, task order dummies) ($\chi^2 = 3.13$, $p = 0.210$) are all jointly insignificant.

Second, we employ other payoff indices from the previous studies that are comparable to RISK and TEMPT: fear, greed, normalized loss and gain (Table A2). We find that greed and normalized gain, which are related to temptation, are highly significant and have a greater impact on cooperation than fear and normalized loss, which are related to risk.

2.5 Discussion and Conclusions

We examine which of risk, temptation, and efficiency is the most influential determinant of cooperation in experimental prisoner's dilemmas. To do this we have subjects play eight one-shot prisoner's dilemma games where payoff parameters vary across games so that indices of risk, temptation, and efficiency vary orthogonally.

We find that the measure of temptation has a significantly negative impact on cooperation. This is the case using our preferred index of temptation, "TEMPT" (Mengel, 2018), but is also the case using alternative measures of temptation ("greed" and "normalized gain"). A majority of subjects are sensitive to the level of temptation: such subjects cooperate more often when the level of temptation is low but defect more often when the level of temptation is high. Cooperation rates also increase with efficiency, though the effect size is much smaller. Cooperation does not vary significantly with risk. Therefore, this finding suggests that curbing the level of temptation is the most important way to improve cooperation in social dilemmas.

This finding supports previous studies (Ahn et al., 2001; Schmidt et al., 2001; Weber, mimeo) that also found increasing temptation significantly reduces cooperation. It also in line with Ahn et al. (2001) which reported a larger impact of greed than fear. Note that our results are similar to Ahn et al. (2001) despite the fact that the levels of RISK and TEMPT used in our experiment are quite different from those used in Ahn et al. (2001) (the levels of (*low* RISK, *high* RISK, *low* TEMPT, *high* TEMPT) are (0.55, 0.90, 0.17, 0.38) in our study compared with (0.17, 0.67, 0.09, 0.29) in Ahn et al. (2001)).

The next question that follows from our findings is: Why do subjects change their decisions with the degree of temptation or efficiency? We exam-

ine whether the effect of payoff indices on cooperation can be explained by using beliefs as a mediating variable. That is switchers might switch to defecting when TEMPT increases because they become more pessimistic about their opponent's cooperativeness, or they might switch to cooperating when EFF increases because they become more optimistic about their opponent's cooperativeness.

Subjects' beliefs do indeed change with the payoff indices: subjects expect others to be more cooperative when efficiency is high or temptation is low. When we examine the effect of indices, controlling for beliefs, we find that TEMPT is still significant. Thus, we conclude that increasing temptation directly reduces willingness to cooperate, independently of expectations of others' behavior.

Overall, our results suggest that both individual preferences over outcomes and beliefs about the opponent's choice affect cooperation in simultaneous prisoner's dilemmas. Further research could separate out the role of preferences and beliefs by studying how cooperation is affected by payoff variations in sequential prisoner's dilemmas, where second-movers directly observe first-movers' decisions and do not need to construct beliefs about their opponent's behavior.

References

Ahn, T., E. Ostrom, D. Schmidt, R. Shupp, and J. Walker. "Cooperation in PD games. Fear, Greed, and History of Play." *Public Choice*, 106, 2001, 137–155.

Arechar, A., S. Gächter, and L. Molleman. "Conducting interactive experiments online" *Experimental Economics*, 21, 2018, 99–131.

Au, W.T., S. Lu, H. Leung, P. Yam and J.M.Y. Fung. "Risk and Prisoner's Dilemma: A Reinterpretation of Coombs' Re-parameterization." *Journal of Behavioral Decision Making*, 25, 2012, 476-490.

Balliet, D., and P.A. Van Lange. "Trust, conflict, and cooperation: A meta-analysis." *Psychological Bulletin*, 139, 2013, 1090-1112.

Charness, G., L. Rigotti, and A. Rustichini. "Social surplus determines cooperation rates in the one-shot Prisoner's Dilemma." *Games and Economic Behavior*, 100, 2016, 113–124.

Cooper, D., D. Dejong, R. Forsythe and T. Ross. "Cooperation without reputation: Experimental evidence from prisoner's dilemma games." *Games and Economic Behavior*, 12(2), 1996, 187–218.

Costa Lobo, M., and J. Curtice. "Personality Politics?: The Role of Leader Evaluations in Democratic Elections." Oxford University Press, 2015.

Croson, R. "Theories of commitment, altruism and reciprocity: Evidence from linear public goods games." *Economic Inquiry*, 45, 2007, 199-216.

Dal Bó, P., and G.R. Fréchette. "On the Determinants of Cooperation in Infinitely Repeated Games: A Survey." *Journal of Economic Literature*, 56(1), 2018, 60–114.

Engel, C., and L. Zhurakhovska. "When is the risk of cooperation worth taking? The prisoner's dilemma as a game of multiple motives." *Applied Economics Letters*, 23:16, 2016, 1157–1161.

Frank, R., T. Gilovich and D. Regan. "Does studying economics inhibit cooperation?." *Journal of Economic Perspectives*, 7(2), 1993, 159–171.

Friedman, D., and R. Oprea. "A continuous dilemma." *American Economic Review*, 102(1), 2012, 337–363.

Gächter, S., D. Nosenzo, E. Renner, and M. Sefton. "Who Makes a Good Leader? Social Preferences and Leading-by-Example." *Economic Inquiry*, 50(4), 2012, 953–967.

Gächter, S., and E. Renner. "Leaders as role models and 'belief managers' in social Dilemmas."*Journal of Economic Behavior and Organization*, 154, 2018, 321–334.

Giamattei, M., K.S. Yahosseini, S. Gachter, and L. Molleman. "LIONESS Lab: a free web-based platform for conducting interactive experiments online." *Journal of Economic Science Association*, forthcoming 2020.

Kroh, M. "Measuring Left-Right Political Orientation: The Choice of Response Format." *Public Opinion Quarterly*, 71(2), 2007, 204–220.

Mengel, F. "Risk and Temptation: A Meta-Study on Prisoner's Dilemma Games." *The Economic Journal*, 128, 2018, 3182–3209.

Moisan, F., R. T. Brincke, R.O. Murphy, and C. Gonzalez. "Not All Prisoner's Dilemma Games Are Equal: Incentives, Social Preferences, and Cooperation." *Decision*, 5(4), 2018, 306–322.

Murnighan, J., and A. Roth. "Expecting continued play in prisoner's dilemma games." *Journal of Conflict Resolution*, 27(2), 1983, 279–300.

Murphy, R.O., and K.A. Ackermann. "Social preferences, positive expectations, and trust based cooperation." *Journal of Mathematical Psychology*, 67, 2015, 45–50.

Neugebauer, T., J. Perote, U. Schmidt, and M. Loos. "Self-biased conditional cooperation: On the decline of cooperation in repeated public goods experiments." *Journal of Economic Psychology*, 30, 2009, 52–60.

Ng, G.T.T., and W.T. Au. "Expectation and cooperation in prisoner's dilemmas: The moderating role of game riskness." *Psychonomic Bulletin and Review*, 23, 2016, 353-360.

Rapoport, A. "A note on the 'index of cooperation' for prisoner's dilemmas." *Journal of Conflict Resolution*, 11(1), 1967, 100-103.

Rapoport, A, and A. Chammah. Prisoner's Dilemma: A Study in Conflict and Cooperation, University of Michigan Press, Ann Arbor, 1965.

Sally, D. "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992." *Rationality and Society*, 7(1), 1995, 58-92.

Schmidt, D., R. Shupp, J. Walker, T.K. Ahn, and E. Ostrom. "Dilemma games: game parameters and matching protocols."*Journal of Economic Behavior and Organization*, 46, 2001, 357–377.

Vlaev, I., and N. Chater. "Game relativity: how context influences strategic decision making." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 2006, 131-149.

Weber, T. "The physiological foundations of social preferences: Do the physiological reactions to social dilemmas vary with SVO?" Mimeo, University of Nottingham.

Appendices

Appendix 2.A Robustness Checks

Table 2.A1 provides regression results using K-index or RISK, TEMPT, and EFF indices, beliefs, task characteristics, individual characteristics and session dummies as explanatory variables. Table 2.A2 provides regression results using alternative payoff indices to RISK and TEMPT.

Table 2.A1: Determinants of Cooperative Choice in Prisoner's Dilemmas: K-index, RISK, TEMPT, and EFF

	(1) All	(2) All	(3) Switcher	(4) Switcher
K-index	0.234***		0.331***	
	(0.087)		(0.116)	
RISK		-0.061		-0.099
		(0.065)		(0.088)
TEMPT		-0.468***		-0.643***
		(0.117)		(0.152)
EFF		0.095		0.141
		(0.075)		(0.101)
Round	-0.009	-0.011*	-0.014	-0.017*
	(0.007)	(0.007)	(0.009)	(0.009)
Belief	0.425***	0.417***	0.472***	0.460***
	(0.071)	(0.070)	(0.079)	(0.079)

	(1) All	(2) All	(3) Switcher	(4) Switcher
B = Coop.	0.052	0.050	0.045	0.041
	(0.035)	(0.035)	(0.045)	(0.045)
BeliefThenChoice	-0.034	-0.030	-0.036	-0.032
	(0.028)	(0.027)	(0.038)	(0.037)
age	0.005	0.005	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
female	0.059	0.059	0.094*	0.094*
	(0.057)	(0.058)	(0.051)	(0.051)
politics	0.016*	0.015*	0.007	0.007
	(0.009)	(0.009)	(0.009)	(0.009)
Mturk exp.	0.010	0.010	-0.026	-0.026
	(0.062)	(0.062)	(0.057)	(0.057)
full-time	-0.009	-0.008	-0.002	-0.000
	(0.062)	(0.062)	(0.057)	(0.057)
higher	-0.052	-0.052	-0.015	-0.015
education	(0.063)	(0.063)	(0.056)	(0.057)
income	-0.014	-0.014	-0.003	-0.003
$\geq$ \$50,000	(0.068)	(0.068)	(0.064)	(0.064)
		Ethnicity- baseline: White		
Asian	0.031	0.031	0.032	0.031
	(0.098)	(0.098)	(0.066)	(0.065)
Black	0.056	0.054	0.017	0.015
or African	(0.083)	(0.082)	(0.098)	(0.098)
Latin	-0.025	-0.027	0.100	0.098
American	(0.137)	(0.136)	(0.111)	(0.111)
		Session - baseline: Session 1		
Session 2	0.160*	0.160*	0.072	0.071
	(0.094)	(0.094)	(0.080)	(0.081)

	(1) All	(2) All	(3) Switcher	(4) Switcher
Session 3	-0.032	-0.033	-0.051	-0.052
	(0.076)	(0.076)	(0.077)	(0.077)
Session 4	0.052	0.054	-0.019	-0.020
	(0.080)	(0.080)	(0.075)	(0.075)
Session 5	-0.003	-0.000	-0.057	-0.056
	(0.085)	(0.084)	(0.084)	(0.083)
Observations	912	912	640	640
Log-likelihood	-470.4	-466.5	-396.1	-392.0
BIC	1,084.0	1,089.7	927.8	932.6
Pseudo R^2	0.091	0.099	0.082	0.092

Notes: Average marginal effects from random effect probit regression with standard errors clustered on individuals. K-index: $\frac{R-P}{T-S}$, RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$. B=Coop. is an indicator variable for whether the cooperative action was labelled as option B. BeliefThenChoice indicates the belief elication is placed in the upper section and the decision task in the lower section of the screen. Mturk.exp indicates the subject has taken part in previous MTurk experiments more than "Once or Twice". Education (Less than a high school degree, High school Diploma, Vocational training, Attended College, Undergraduate, Graduate school, Doctoral or further), Ethnicity (American Indian or Alaska Native, Asian, Black of American African, Native Hawaiian or Other Pacific Islander, White, Latin American, Others), Employment (Full-time worker, Part-time worker, Unemployed, Student, Retired, Self-employed, Homemaker, Unable to work), Income (Less than \$29,999, \$30,000 $\sim$ \$49,999, \$50,000 $\sim$ \$69,999, \$70,000 $\sim$ \$89,999, \$90,000 or more) Politics (0 to 10, 0 means the most left and 10 means the most right). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 2.A2: Determinants of Cooperative Choice in Prisoner's Dilemmas: Fear, Greed, Normalized Loss, and Normalized Gain

	(1) All	(2) All	(3) Switcher	(4) Switcher
Fear	-0.141		-0.208	
	(0.095)		(0.129)	
Greed	-0.457***		-0.632***	
	(0.113)		(0.149)	
Normalized loss		0.005		0.005
		(0.013)		(0.017)
Normalized gain		-0.041***		-0.056***
		(0.015)		(0.021)
Round	-0.011*	-0.010	-0.016*	-0.016*
	(0.007)	(0.007)	(0.009)	(0.009)
Belief	0.417***	0.427***	0.460***	0.473***
	(0.070)	(0.070)	(0.078)	(0.079)
B = Coop.	0.052	0.052	0.044	0.044
	(0.034)	(0.034)	(0.045)	(0.044)
BeliefThenChoice	-0.033	-0.033	-0.036	-0.035
	(0.027)	(0.028)	(0.037)	(0.038)
age	0.005	0.005	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
female	0.059	0.059	0.094*	0.094*
	(0.058)	(0.057)	(0.051)	(0.051)
politics	0.015*	0.016*	0.007	0.007
	(0.009)	(0.009)	(0.009)	(0.009)
Mturk exp.	0.009	0.010	-0.027	-0.026
	(0.062)	(0.062)	(0.057)	(0.057)
full-time	-0.009	-0.009	-0.001	-0.001
	(0.062)	(0.062)	(0.057)	(0.057)

	(1) All	(2) All	(3) Switcher	(4) Switcher
higher	-0.051	-0.053	-0.015	-0.016
education	(0.063)	(0.063)	(0.057)	(0.056)
income	-0.014	-0.013	-0.003	-0.003
$\geq$ \$50,000	(0.068)	(0.068)	(0.064)	(0.064)
		Ethnicity- baseline: White		
Asian	0.030	0.032	0.030	0.033
	(0.098)	(0.098)	(0.066)	(0.066)
Black	0.055	0.055	0.016	0.016
or African	(0.083)	(0.083)	(0.099)	(0.099)
Latin	-0.025	-0.025	0.100	0.099
American	(0.136)	(0.136)	(0.111)	(0.110)
		Session - baseline: Session 1		
Session 2	0.159*	0.159*	0.071	0.072
	(0.094)	(0.094)	(0.081)	(0.080)
Session 3	-0.034	-0.032	-0.053	-0.051
	(0.077)	(0.076)	(0.077)	(0.077)
Session 4	0.054	0.052	-0.019	-0.020
	(0.080)	(0.079)	(0.075)	(0.075)
Session 5	-0.001	-0.003	-0.056	-0.057
	(0.085)	(0.085)	(0.084)	(0.084)
Observations	912	912	640	640
Log-likelihood	-467.2	-470.6	-392.9	-396.4
BIC	1,084.4	1,091.2	928.0	934.9
Pseudo R^2	0.097	0.091	0.090	0.082

Notes: Average marginal effects from random effect probit regression with standard errors clustered on individuals. Fear: $\frac{P-S}{T-S}$, Greed: $\frac{T-R}{T-S}$, Normalized loss: $\frac{P-S}{R-P}$, Normalized gain: $\frac{T-R}{R-P}$. B=Coop. is an indicator variable for whether the cooperative action was labelled as option B. BeliefThenChoice indicates the belief elication is placed in the upper section and the decision task in the lower section of the screen. Mturk.exp indicates the subject has taken part in previous MTurk experi-

ments more than "Once or Twice". Education (Less than a high school degree, High school Diploma, Vocational training, Attended College, Undergraduate, Graduate school, Doctoral or further), Ethnicity (American Indian or Alaska Native, Asian, Black of American African, Native Hawaiian or Other Pacific Islander, White, Latin American, Others), Employment (Full-time worker, Part-time worker, Unemployed, Student, Retired, Self-employed, Homemaker, Unable to work), Income (Less than \$29,999, \$30,000 $\sim$ \$49,999, \$50,000 $\sim$ \$69,999, \$70,000 $\sim$ \$89,999, \$90,000 or more) Politics (0 to 10, 0 means the most left and 10 means the most right). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Appendix 2.B Experimental Instructions

Welcome

Thank you for accepting this HIT. To complete this HIT, you must make some decisions. Including the time for reading these instructions, the HIT will take about 30 minutes to complete. If you are using a desktop or laptop to complete this HIT, we recommend that you maximize your browser screen (press F11) before you start.

It is important that you complete this HIT without interruptions. During the HIT, please **do not close this window or get distracted from the task.** If you close your browser or leave the task, you will not be able to re-enter and we will not be able to pay you.

In this HIT, you will be matched with one other participant. Each of you will make decisions for 8 decision situations. In each situation, each of you will earn Tokens depending on your decisions.

At the end of the HIT, one of the decision situations will be randomly chosen. Your earnings from this situation will be converted from Tokens to Dollars at a rate of 1**00 Tokens = $ 1.** This will be added to **your participation fee of $1.00.** Depending on your decisions, you may make up to $8.00 more in addition to the $1.00 participation fee. In the same way, Tokens earned by the person matched with you in that same situation will also be converted to Dollars at a rate of 100 Tokens = $ 1.

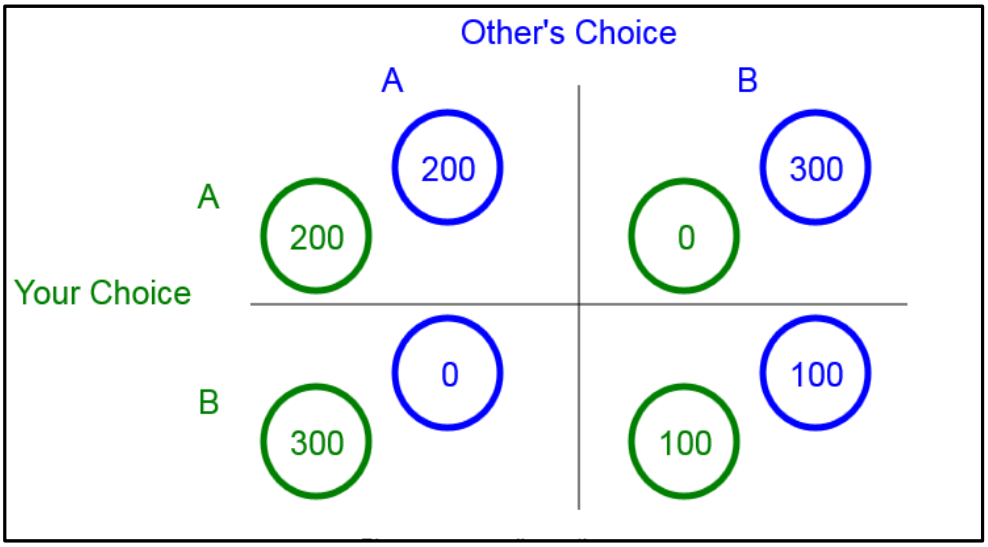
You will receive a code to collect your payment via MTurk upon completion.

Please click "Continue" to start the HIT.

INSTRUCTIONS

The HIT consists of 8 decision situations.

Each decision situation will be presented on a screen like the **example screen** below.



You and the other person will be making choices between **A** and **B**. Your earnings are the values in the green circle, and the other person's earnings are the values in the blue circle. The table is read as follows:

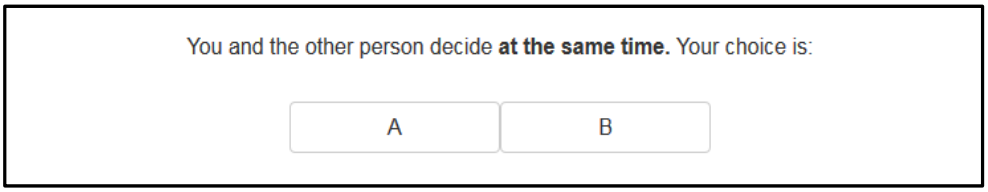
- If you choose A and the other person chooses A, you will earn 200 Tokens and the other person will earn 200 Tokens.
- If you choose A and the other person chooses B, you will earn 0 Tokens and the other person will earn 300 Tokens.
- If you choose B and the other person chooses A, you will earn 300 Tokens and the other person will earn 0 Tokens.
- If you choose B and the other person chooses B, you will earn 100 Tokens and the other person will earn 100 Tokens.

Please note that the values in the table will differ in each decision situation.

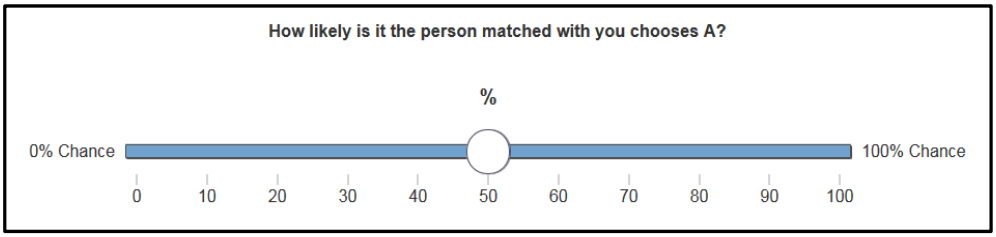
Tasks

In each decision situation, you must complete **two types** of tasks, which we will refer to below as the "decision" and "prediction".

● For the "decision" task, you will see the following screen and you must choose A or B:



● For the "prediction" task, you will see the following screen and you must indicate how likely you think it is that the other person will choose A:



During the HIT, you will not receive any feedback on the other person's choice or the outcomes of the decision situations.

Your dollar earnings

On completion of the HIT, you will be paid your participation fee of $ 1.

In addition, one of the decision situations will be randomly chosen for your additional dollar earnings. Your earnings and the other person's earnings will be determined depending on choices of you and the other person in that situation. Two examples should make this clear.

Example 1. Assume that you choose A and the other person matched with you chooses A in the above example screen. As a consequence, you will earn 200 Tokens and the other person will earn 200 Tokens.

Example 2. Assume that you choose B and the other person matched with you chooses A in the above example screen. As a consequence, you will earn 300 Tokens and the other person will earn 0 Tokens.

At the end of the HIT

On completion of the HIT, one of the decision situations will be randomly chosen as explained above. You will be informed of your choices and earnings for that decision situation, and you will be paid these earnings in addition to your participation fee.

Note that we will not be able to pay you if you do not complete the HIT. If the person you are matched with does not complete the HIT, the computer will randomly select one of the four possible earnings in the randomly chosen decision situation, and you will be paid these earnings in addition to your participation fee.

Your participation fee and the additional earnings will be paid to you within two working days.

Chapter 3

The Variability of Conditional Cooperation in Sequential Prisoner's Dilemmas

3.1 Introduction

Conditional cooperation is widely observed in social dilemmas. Whereas the pursuit of narrowly-defined selfish interests would result in a lack of cooperation, many people are willing to forgo their selfish interests and cooperate, but only as long as other group members also cooperate. This pattern of behavior is particularly clear in controlled experiments investigating contributions to public goods (Brandts and Schram, 2001; Fischbacher, Gächter and Fehr, 2001; Gächter, 2007; Keser and van Winden, 2000; Kocher et al., 2008). These experiments also reveal substantial heterogeneity: for example, in some of these studies some group members are classified as "free riders" (i.e. defecting regardless of the behavior of others), others as "conditional cooperators" (i.e. cooperating as long as others do so), and still others as "unconditional cooperators" (i.e. cooperating independently of the behavior of others). Not

much is known, however, about whether such a classification reflects stable personality traits whereby the participant would exhibit similar behavioral patterns in similar situations, or whether the classification applies only to the specific experimental setting and parameters.

In this chapter, we examine whether the behavioral pattern exhibited by a given participant, such as conditional cooperation, varies across payoff variations. Two previous studies have examined the variability of conditional cooperation across payoff variations. In a meta-study, Thöni and Volk (2018) found that the proportion of conditional cooperators is similar across 17 public goods experimental studies, employing different parameters (e.g., marginal per capita return, group size). In contrast, Clark and Sefton (2001), using a between-subject design where different subjects are assigned to different payoff treatments, find that conditional cooperation in sequential prisoner's dilemmas significantly decreases when the payoff to defecting against a cooperator doubles. In contrast to these studies, we examine the individual-level variability of conditional cooperation using a within-subject design: the experiment consists of eight games where payoffs differ, and subjects make decisions in all games.

Examining the within-subject variability of conditional cooperation across payoff variations is important for at least two reasons. First, it allows us to understand the nature of conditional cooperation: whether conditional cooperation reflects underlying social preferences, or whether conditional cooperation reflects a desire to reciprocate the cooperation of others in a way that is robust to changes in material incentives.[1] Social preference models, which define pref-

[1] As discussed by Fehr and Fischbacher (2004), Gächter et al. (2017), and Katuscak and Miklánek (2018), conformity to what is perceived as "socially appropriate" and willing to sacrifice material payoffs in order to follow such norms could also be a candidate explanation for conditional cooperation.

erences over one's own and other's material payoffs (e.g., Andreoni and Miller, 2002; Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Cox, Friedman and Gjerstad, 2007; Fehr and Schmidt, 1999), are capable of explaining conditional cooperation, but at the same time predict that it will be influenced by material incentives. In contrast, if conditional cooperation reflects a principled stand against free-riding, eschewing material gains in order to reciprocate the cooperation of others, then conditional cooperation is expected to be robust across payoff variations.

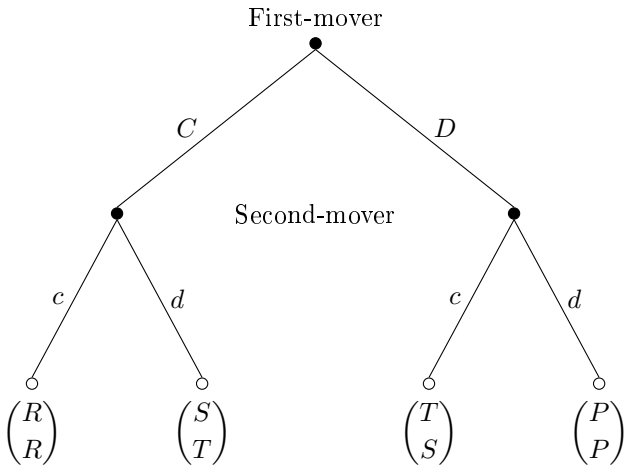
Second, the efficacy of interventions to promote cooperation depends on whether conditional cooperation is influenced by payoff variations. For example, leading by example would be an effective mechanism to achieve cooperative outcomes if followers are generally conditionally cooperative (Gächter, Nosenzo, Renner and Sefton, 2012). On the other hand, if conditional cooperation is sensitive to payoffs, then this implies that there would be settings where leading by example is ineffective.

In order to study the within-subject variability of conditional cooperation, we use the sequential prisoner's dilemma. In the sequential prisoner's dilemma (henceforth SPD), there are two players - the first-mover and the second-mover - who sequentially choose whether to cooperate or defect. The extensive form of the SPD is shown in Figure 3.1.

The first mover chooses to either cooperate (C) or defect (D), and, after observing this choice, the second-mover then chooses to cooperate (c) or defect (d). Subjects' earnings are maximized when both cooperate and each player receives R. However, if the first-mover cooperates, the second-mover maximizes her own earnings by defecting, in which case the first-mover receives S and the second-mover receives T. If the first-mover defects, the second-mover

maximizes her own earnings by also defecting, so that each player receives P. If players are own earnings maximizers, and this is common knowledge, there is a unique equilibrium where the first-mover defects and the second-mover defects regardless of the first-mover's choice.

Figure 3.1: The Sequential Prisoner's Dilemma (SPD)



Note: $T > R > P > S$ and $2R > T + S$

Previous experiments with the SPD have shown that second-movers hardly ever cooperate in response to first-mover defection (Clark and Sefton, 2001; Miettinen, Kosfeld, Fehr and Weibull, 2020). In contrast, the response to first-mover cooperation is less predictable, and there are non-negligible proportions of subjects who cooperate (i.e., conditional cooperators) and defect (i.e., free-riders). There are two factors related to payoffs that could plausibly affect the response to cooperation in the SPD. First, *damage* refers to the cost imposed on the first-mover when the second-mover responds by defecting: $R - S$. Second, *gain* refers to the gains to the second-mover from defecting in response to cooperation: $T - R$. We define DAMAGE ($\frac{R-S}{R}$) as the percentage loss imposed on the first-mover by defecting in response to cooperation, while we define GAIN ($\frac{T-R}{R}$) as the percentage gain from defecting against a cooperator.

Compared to the public goods game which has frequently been used to study conditional cooperation, the SPD is simpler in that it has only two players and binary actions. Moreover, it allows us to separately vary the factors of damage and gain.[2] Using eight SPDs where payoffs vary, we implement a within-subject experiment in which every subject makes decisions in the role of both first-mover and second-mover for each game. For second-mover decisions, we ask how the subject would respond to defect and how they would respond to cooperate, with the actual decision being determined by the response to the first-mover's actual choice. Thus, we elicit the second-mover's strategy for playing each SPD. To rule out confounding factors, such as belief updating, no feedback on any of the individual games is provided until the end of the experiment. Our experimental design allows us to examine the following two questions: Is conditional cooperation influenced by payoff changes? If so, how conditional cooperation is influenced by the factors of damage and gain?

Our main findings are two-fold. First, we find that conditional cooperation is sensitive to payoff variations. Only 26% of subjects submitted the same second-mover strategy across all eight games. 13% of subjects were consistently free-riders and 13% of subjects were consistently conditional cooperators. The remaining 74% of subjects changed their second-mover strategy at least once across games.

Second, conditional cooperation is significantly influenced by DAMAGE and GAIN. Conditional cooperation increases with DAMAGE: second-movers are more likely to conditionally cooperate when free-riding has a larger negative impact on the first-mover's earnings. In contrast, conditional cooperation

[2] In the standard public goods game (see Chaudhuri, 2011), varying the MPCR (Marginal Per Capita Return) affects the levels of GAIN and DAMAGE simultaneously, making it difficult to disentangle the effect of GAIN and DAMAGE on decisions.

decreases with GAIN: second-movers are less likely to conditionally cooperate when they gain more from free-riding.

Our finding suggests that classifications of individuals as conditional cooperators or free-riders should not be generalized to other games with different material payoffs. In other words, a conditional cooperator in one game may be a free-rider in another, and vice versa. The variability of conditional cooperation with GAIN and DAMAGE suggests that the effectiveness of leading by example in promoting cooperation may be limited when GAIN is high or DAMAGE is low.

This study is also related to research that uses theories of social preferences to explain unselfish behavior in various contexts. The within-subject variation of conditional cooperation with the levels of damage and gain is consistent with the predictions of several distributional preference models (e.g., Andreoni and Miller, 2002; Fehr and Schmidt, 1999). To this extent, our results support the view that conditional cooperation reflects underlying social preferences.

The remainder of the chapter is organized as follows. In Section 3.2, we review the related literature that examines the variability of conditional cooperation. Section 3.3 presents our experimental design and procedures. Section 3.4 presents our results, and we conclude in Section 3.5.

3.2 Related Literature

In this section, we review related studies that examine the variability of conditional cooperation. A number of studies examine the variability of conditional cooperation over time with mixed results. Brosig et al. (2007) conducted SPDs three times within three months using the same subjects and random-matching, and found that the rate of conditional cooperation diminished across

repetitions. Muller et al. (2008) elicited subjects' strategy across five repetitions of a public goods game. Although, only 37% of subjects always choose the same strategy across all five games, previous choices were useful predictors of subsequent choices. For example, 69% of subjects who conditionally cooperated in any of the first four games also conditionally cooperated in the fifth game. Volk et al. (2012), which elicited subjects' strategies in a public goods game three times over the course of five months, observed that conditional cooperation was remarkably stable over time. Half of their subjects chose the same strategy in all three games, and 71% of these conditionally cooperated.

Two further studies examine the variability of conditional cooperation across different contexts by comparing behavior in a public goods game and a SPD. Eichenseer and Moser (2020) and Mullett et al. (2020) report that subjects who are conditionally cooperative in a SPD are also conditionally cooperative in a public goods game.

We are not aware of any study examining how within-subject variation of payoffs affects conditional cooperation.[3] In fact, we are only aware of two studies that examine whether payoff variation affects conditional cooperation. Thöni and Volk (2018) found that the proportion of conditional cooperators is quite similar across 17 public goods experiments, which employ different parameters (i.e. marginal per capita return, group size). In contrast, Clark and Sefton (2001), using a between-subjects SPD experiment, found that doubling the temptation payoff, T, resulted in a significantly lower rate of conditional cooperation. Our study differs from these two studies in that we employ a within-subject experiment systematically varying damage and gain: using the

[3] Several studies examine how decisions in the simultaneous prisoner's dilemmas are influenced by payoff variations (e.g., Ahn et al., 2001; Gächter, Lee and Sefton, 2020; Schmidt et al., 2001; Vlaev and Chater, 2006).

same subjects, we ask subjects to make decisions in eight sequential prisoner's dilemmas where payoffs vary.[4] Our design allows us to examine how conditional cooperation is influenced by payoff variations, at the individual level.

3.3 Experimental Design & Procedures

3.3.1 Experimental Design

The experiment is based on a simple sequential two-player prisoner's dilemma game. Each subject must decide whether to cooperate or defect. The first-mover decides first and the second-mover decides after observing the first-mover's decision. Subjects had to make decisions in eight such games with varying payoffs.

Table 3.1 shows the payoff parameterization used in the experiment. This is the same parameterization used in Gächter, Lee and Sefton (2020) that studies cooperation in simultaneous PDs. Payoffs were chosen to be strictly positive multiples of ten in order to avoid zero or non-rounded payoffs. R (500) is constant across all games, while there are two distinct values of T (600, 800) and P (200, 400), and four distinct values of S (20, 90, 40, 180). These eight games were chosen to orthogonally vary risk ($\frac{P-S}{P}$), temptation ($\frac{T-R}{T}$), and efficiency ($\frac{R-P}{R}$) to examine how these factors affect play in simultaneous prisoner's dilemmas.

[4] We note two further, potentially important, differences between their design and ours. In their Double Temptation treatment, the payoffs are such that mutual cooperation does not maximize combined earnings. Second, within a treatment their subjects play ten repetitions of the game (using a perfect stranger matching protocol), giving feedback at the end of each game.

Table 3.1: Payoff Parameterizations for Sequential Prisoner's Dilemmas

Game	T	R	P	S	RISK	TEMPT	EFF	DAMAGE	GAIN
G1	600	500	200	90	0.55	0.17	0.60	0.82	0.20
G2	600	500	200	20	0.90	0.17	0.60	0.96	0.20
G3	800	500	200	90	0.55	0.38	0.60	0.82	0.60
G4	800	500	200	20	0.90	0.38	0.60	0.96	0.60
G5	600	500	400	180	0.55	0.17	0.20	0.64	0.20
G6	600	500	400	40	0.90	0.17	0.20	0.92	0.20
G7	800	500	400	180	0.55	0.38	0.20	0.64	0.60
G8	800	500	400	40	0.90	0.38	0.20	0.92	0.60

Note: RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$

These indices capture important motivations that are relevant in a simultaneous PD, where players face uncertainty about the choice of the opponent (see Gächter et al, 2020, for a discussion). In the sequential PD, we are mainly interested in the second-mover's response to cooperation, and so we focus on the indices of DAMAGE and GAIN.[5] DAMAGE ($\frac{R-S}{R}$) refers to the losses imposed on the first-mover when the second-mover defects against cooperation, while GAIN ($\frac{T-R}{R}$) represents the gains to the second-mover from defecting against to the first-mover's cooperation.

The implied values of DAMAGE and GAIN are as follows. The value 0.64 (0.96) of DAMAGE implies that an individual who defects against a cooperator will reduce the first-mover's monetary payoff by 64% (96%) compared to the payoff from mutual cooperation. The value 0.20 (0.60) of GAIN implies that an individual who defects against a cooperator will accrue a 20% (60%) gain

[5] Note that DAMAGE is more closely related to the distributional outcomes determined by the second-mover than RISK. RISK is not directly relevant to the second-mover's response to cooperation because P is the outcome of mutual defection.

in monetary payoff compared to the payoff from mutual cooperation. Note that in this parameterization there are two levels of efficiency, and for each of these we study a 2×2 variation in DAMAGE and GAIN.

Our main interest lies in exploring how payoff variations - especially with DAMAGE and GAIN - affect the within-subject variability of conditional cooperation. The Nash equilibrium assuming self-interested agents is mutual defection, regardless of the levels of DAMAGE and GAIN. Therefore, the self-interested agent model predicts that conditional cooperation never occurs in SPDs. Distributional preference models (e.g., Andreoni and Miller, 2002; Fehr and Schmidt, 1999), which define an individual's preferences over own and other's material earnings, allow for the possibility of rational conditional cooperation for some payoff parameters. These models predict that second-movers are more likely to conditionally cooperate as T decreases, S decreases, or R increases: This implies that the likelihood of conditional cooperation increases with DAMAGE and decreases with GAIN.[6]

3.3.2 Experimental Procedures

We conducted the online interactive experiment in Spring 2019 using MTurk. Subjects were residents of the United States. We conducted five sessions with a total of 138 participants. None of the subjects participated in more than one session. Each participant was paired with another subject after he/she had read the instructions and passed the control questions. Each pair then played all eight games of Table 3.1 with no feedback.

Subjects completed two types of decisions in each game: First-mover's

[6] In Appendix A at the end of the chapter, we provide the theoretical predictions for second-mover's choice using the following distributional preference models: Andreoni and Miller (2002), and Fehr and Schmidt (1999).

decision and Second-mover's decision. Both decision tasks were presented on the same screen. In the *First-mover's decision*, they simply chose whether to cooperate or to defect as a first-mover. In the *Second-mover's decision*, we asked subjects to decide in the following two situations: i) if the first-mover cooperates, and ii) if the first-mover defects. Therefore, we elicited second-mover's decision using the strategy method (Selten, 1967).[7] Rather than use the terms "cooperate" or "defect", we labeled options neutrally as A or B, with labeling randomly chosen at the pair level in each game. To control for potential order effects, we randomized the sequence of games and the order of tasks (first-mover's decision and second-mover's decision) at the pair level.

For each game, subjects had to answer eight additional control questions about the payoffs before making decisions. These additional control questions were intended to ensure that subjects understood the implications of their decisions and recognized the payoff changes across games.

Subjects did not receive any feedback on the other's choice or the outcome of each game until the end of the experiment. Once subjects completed the tasks for all games, we asked them to complete a short post-experimental questionnaire eliciting basic demographic information. This included a self-assessment of subjects' political orientation, using an 11-point left-right scale that is frequently used in surveys (e.g. Costa Lobo and Curtice, 2015; Kroh, 2007).

We implemented the experiment using the software LIONESS (Giamattei, Yahosseini, Gächter and Molleman, 2020). Subjects were paired with another participant on a real-time basis and they conducted each task at the same

[7] Regarding potential differences between responses elicited using the strategy method and those using a direct response method, previous studies found no statistical differences in subjects' responses between these two methods (see Brandts and Charness, 2000; 2011 for a review).

time. This implies a subject had to wait until the opponent made a decision to proceed to the next game. As subjects needed to wait until their opponent made a decision, long waiting times could increase the risk of reduced attention. We took the following measures to retain attention and encourage successful completion of the experiment. Before participants entered the experiment, we told them to avoid distractions during the experiment. In addition, participants who were inactive for more than 30 seconds (i.e. no mouse movement or no keyboard input) got an alert voice message and a blinking text on their browser. If an inactive participant did not respond to the alert message for a further 30 seconds, such an inactive participant was removed from the experiment and the remaining person was able to continue the experiment.

To elicit subjects' responses in an incentive-compatible way, we implement the following payment scheme. At the end of the session, one out of eight games was randomly chosen for payment. If both subjects completed the entire experiment, they were paid according to the outcome of the randomly chosen game as follows. One of the pair was randomly chosen to be the first-mover, and the other was selected to be the second-mover. Then, subjects were reminded of their decisions and informed about the outcome for the randomly chosen game. For the second mover's decision we used their conditional response to the first-mover's decision. If one of the pair had dropped out during the experiment, the computer randomly selected the payoff-relevant game for the remaining subject. Then the computer randomly selected one out of four monetary outcomes (i.e. T, R, P, or S) of the chosen game for payment to the remaining subject. We explained this payment scheme clearly in the instructions. This payment procedure gives subjects a monetary incentive to take both first-mover's decision and second-mover's decisions seriously in all

games as all those decisions can become payoff-relevant.

As normally occurs in online experiments, there was a non-negligible attrition rate: 32 out of 138 subjects (23%) dropped out during the experiment.[8] For subjects who completed the experiment, the average age was 34 years (between 19 and 65 years) and 37% were female. Subjects' earnings ranged from $ 1.20 to $9.00, averaging $ 4.59. On average, the experiment lasted about 30 minutes, including the completion of a post-experimental questionnaire. Subjects were informed of their payment immediately upon completion of the experiment and were paid within 24 hours.

3.4 Results

Our focus is on second-mover decisions. Since the second-mover can condition choices on the first-mover's decision, the second-mover's decision gives us a direct measure of conditional cooperation.[9]

There exist four different strategies for the second-mover's decision. Conditional cooperation (henceforth CC) involves cooperating if and only if the first-mover cooperates. Free-riding (FR) involves defecting regardless of the first-mover's choice. Unconditional cooperation (UC) involves cooperating regardless of the first-mover's choice. Lastly, mis-matching (MM) occurs when the second-mover defects in response to cooperation and cooperates in response to defect.

As in most social dilemma experiments (Fischbacher et al., 2001; Fallucchi et al., 2019; Muller et al., 2008), we find that CC and FR strategies predom-

[8] The dropout rate in our experiment is not too different from that of similar interactive online experiments. For example, Arechar et al. (2018) report a 20% dropout rate in their interactive 4-player public goods game, and Gächter et al. (2020) reports 24% dropout rate in their interactive eight simultaneous prisoner's dilemma game.

[9] For completeness, we report first-mover decisions in Appendix B.

inate: the average proportion of CC and FR in eight games is 38% and 45%, respectively. In contrast, the average proportion of UC and MM in eight games is only 12% and 5%, respectively. Thus we focus on CC and FR strategies.

The following analysis is structured to discuss our main research questions. First, does conditional cooperation vary with payoffs? If so, are there systematic patterns between within-subject variations in conditional cooperation and payoff variations? For our analysis, we only include the decisions of subjects who completed the experiment: thus our data set consists of 848 observations (106 subjects $\times$ eight games).

3.4.1 Does Conditional Cooperation Vary with Payoffs?

Only 26% of subjects use the same strategy in all eight games: 13% always defect, 13% always conditionally cooperate. None of the subjects always unconditionally cooperate or always choose mismatching strategy. The remaining 74% of subjects change strategies across games: we refer to these subjects as "switchers" for the rest of the analysis. On average, switchers conditionally cooperate three times out of eight games. Similarly, switchers free-ride three times out of eight games, on average. The finding that a majority of subjects change their second-mover's decisions across games suggests that variation in DAMAGE and GAIN may lead to the variations of conditional cooperation.

On aggregate, across all eight games, 38% of strategies are conditional cooperation. Table 3.2 reports the proportion of CC strategies depending on the levels of DAMAGE and GAIN in the *high* and *low* efficiency games. We find that the proportion of CC strategies exhibits a clear relationship with DAMAGE and GAIN: the proportion of CC strategies increases with DAMAGE and decreases with GAIN. In the *high* DAMAGE games, conditional cooper-

ation increases by 5.9 percentage points on average (the absolute differences range from 3.8 to 7.5 percentage points). In the *high* GAIN games, conditional cooperation decreases by 9.2 percentage points on average (the absolute differences range from 2.8 to 16.9 percentage points).

Table 3.2: The Proportion of Conditional Cooperation Strategies

(a) *high* efficiency games ($P = 200$)

		DAMAGE		
		Low (= 0.82)	*High* (= 0.96)	Δ (*H*-*L*)
GAIN	*Low* (= 0.20)	35.8%	42.5%	+6.7%p
	High (= 0.60)	33.0%	38.7%	+5.7%p
	Δ (*H*-*L*)	−2.8%p	−3.8%p	

(b) *low* efficiency games ($P = 400$)

		DAMAGE		
		Low (= 0.64)	*High* (= 0.92)	Δ (*H*-*L*)
GAIN	*Low* (= 0.20)	44.3%	48.1%	+3.8%p
	High (= 0.60)	27.4%	34.9%	+7.5%p
	Δ (*H*-*L*)	−16.9%p***	−13.2%p***	

Note: DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$ indicate p - values based on Mcnemar's test.

To examine whether conditional cooperation varies significantly across games, we conduct pairwise comparisons using Mcnemar's test. We compare the games one by one holding the level of efficiency (i.e. EFF) constant while varying each payoff index (i.e. DAMAGE, GAIN). In the *low* efficiency games (e.g., G5 $\sim$ G8), there exist significant differences depending on the levels of GAIN: subjects are less likely to conditionally cooperate when GAIN increases (G5 vs G7: Mcnemar's test: $\chi^2 = 10.80$, $p = 0.001$, G6 vs G8: $\chi^2 = 7.00$, $p = 0.008$). For DAMAGE, the differences are not statistically significant al-

though subjects are more likely to conditionally cooperate when DAMAGE increases. In the *high* efficiency games (e.g., G1 $\sim$ G4), the overall pattern of conditional cooperation with DAMAGE and GAIN are similar to the *low* efficiency games: subjects are more likely to conditionally cooperate with increasing DAMAGE or decreasing GAIN. However, the differences in the *high* efficiency games are relatively small and insignificant.

Overall, there is clear evidence that conditional cooperation decreases with GAIN. While the pairwise comparisons of the effect of DAMAGE are not significant, the consistent finding from four comparisons (corresponding to 2×2 combinations of EFF and DAMAGE) that conditional cooperation decreases with GAIN is suggestive. In the next section, we use regression methods to analyse the relationship between conditional cooperation and payoff factors.

3.4.2 Regression Analysis

In this subsection, we examine how the second-mover's decisions are influenced by payoff variations using regression analysis. Table 3.3 shows the results of panel multinomial logit regressions: Column (1) examines how conditional cooperation varies with the indices of DAMAGE ($\frac{R-S}{R}$), GAIN ($\frac{T-R}{R}$), and EFF ($\frac{R-P}{R}$). Column (2) provides corresponding regression results for the subsample of switchers. In columns (3) and (4), we include controls for individual characteristics and session effects in the regressions for all subjects and the subsample of switchers, respectively. In all regressions, we include a control for the order in which subjects played the games (Round $\in$ {1, 2, ..., 8} is the position of the game in the sequence of eight games).

First of all, we find that DAMAGE and GAIN have a significant effect on conditional cooperation (Column (1)): conditional cooperation increases

with DAMAGE but decreases with GAIN. Second-movers are more likely to conditionally cooperate when free-riding has a larger negative impact on the first-mover's earnings: in our experiment, the probability of CC increases by 2.7 percentage points when DAMAGE increases by 0.1 point. For GAIN, second-movers are less likely to conditionally cooperate when they gain more from free-riding: the probability of CC decreases by 2.5 percentage points when GAIN increases by 0.1 point.

Second, switchers are more sensitive to payoff variations (Column (2)). For switchers, who represent 74% of subjects, the effect sizes of DAMAGE (0.356) and GAIN (-0.323) are approximately 1.3 times greater than the effect sizes based on the complete sample. The overall patterns of conditional cooperation for the subsample of switchers are similar to those patterns based on the complete sample, but the effect sizes of payoff indices become larger for the switchers.

Third, the results are robust to a set of controls for individual characteristics and session effects (Columns (3) and (4)). The significance level of regressors are unchanged, and the effect sizes of damage and gain are similar to the results that do not control individual and task characteristics and session dummies. Except for age, individual characteristics (i.e. gender, political orientation, MTurk experience, employment, education, income level, ethnicity) (Wald test, $\chi^2 = 25.01$, $p = 0.247$) and task characteristics (labeling of cooperative choice as A or B, task order dummies) ($\chi^2 = 8.03$, $p = 0.236$) are all jointly insignificant. Session effects are marginally significant at the 10% level ($\chi^2 = 18.65$, $p = 0.097$).

We do not find a significant relation between the index of efficiency (EFF) and conditional cooperation: in all regressions the effect size of EFF is small

and insignificant. Lastly, we also observe a round effect: the probability of CC decreases by 1.6 - 2.2 percentage points from one game to the next.

Table 3.3: Determinants of Conditional Cooperation

	(1) All	(2) Switcher	(3) All	(4) Switcher
DAMAGE	0.274**	0.356**	0.276**	0.374**
	(0.113)	(0.147)	(0.113)	(0.150)
GAIN	-0.245***	-0.323***	-0.263***	-0.348***
	(0.067)	(0.087)	(0.067)	(0.087)
EFF	-0.101	-0.133	-0.096	-0.127
	(0.073)	(0.096)	(0.076)	(0.101)
Round	-0.016**	-0.020**	-0.017**	-0.022**
	(0.007)	(0.009)	(0.007)	(0.009)
Controls	*No*	*No*	*Yes*	*Yes*
Observations	848	624	808	600
Log-likelihood	-731.4	-656.0	-676.8	-610.6
BIC	1,604.4	1,447.2	1775.5	1624.2

Note: Average marginal effects from panel multinomial logit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$. Controls: the variables of demographic information, task characteristics, and session effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Note that free-riding is also prevalent in our experiment as it represents 45% of second-mover strategies. Thus, Table 3.4 reports the determinants of free-riding based on the same panel multinomial logit regressions.

We find that whether to free-ride is also significantly influenced by the variations of DAMAGE and GAIN: free-riding decreases with DAMAGE, but increases with GAIN. Second-movers are more likely to free-ride when free-riding has a smaller negative impact on the first-mover's earnings (decreasing DAMAGE). In contrast, second-movers are more likely to free-ride when the gains from free-riding become larger (increasing GAIN). Note that the signs of the marginal effects on FR are exactly opposite to the marginal effects on CC. This suggests that subjects switch from "conditional cooperators" to "free-riders" or vice versa when the levels of DAMAGE and GAIN vary.

Table 3.4: Determinants of Free Riding

	(1) All	(2) Switcher	(3) All	(4) Switcher
DAMAGE	-0.260**	-0.337**	-0.227*	-0.308*
	(0.118)	(0.157)	(0.118)	(0.160)
GAIN	0.155**	0.200**	0.160**	0.211**
	(0.069)	(0.092)	(0.068)	(0.093)
EFF	0.008	0.007	-0.012	-0.017
	(0.070)	(0.093)	(0.070)	(0.097)
Round	0.021***	0.028***	0.021***	0.029***
	(0.006)	(0.008)	(0.006)	(0.008)
Controls	*No*	*No*	*Yes*	*Yes*
Observations	848	624	808	600
Log-likelihood	-731.4	-656.0	-676.8	-610.6
BIC	1,604.4	1,447.2	1775.5	1624.2

Note: Average marginal effects from panel multinomial logit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$. Controls: the variables of demographic information, task characteristics, and session effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 3.4 reports the following findings, which are similar to the results for the determinants of conditional cooperation (Table 3.3). First, switchers are more sensitive to variations of DAMAGE and GAIN: the marginal effects of DAMAGE and GAIN for switchers are 1.3 times greater than the marginal effects for all subjects. Second, EFF has an insignificant effect on free-riding. Third, there exists a round effect: the probability of free-riding increases by 2.1 - 2.9 percentage points from one game to the next.

Including controls for individuals and session effects (Columns (3)-(4)) weakens the significance of DAMAGE, but other than this, the overall results are qualitatively similar when individual characteristics and session dummies are included: the significance of GAIN is not affected by the inclusion of control variables, and the effect sizes of DAMAGE and GAIN do not change much with the inclusion of control variables.[10]

[10] We also examine how unconditional cooperation and mismatching are influenced by payoff variations. Neither are significantly affected by DAMAGE, GAIN, or EFF.

3.5 Discussion and Conclusions

We examine the within-subject variability of conditional cooperation across payoff variations. To do this, we have subjects play eight one-shot sequential prisoner's dilemma games with varying payoff parameters.

We find that conditional cooperation varies across payoff variations. Most subjects neither always conditionally cooperate nor always free-ride. Instead, the within-subject variation of conditional cooperation has a clear pattern with the levels of damage and gain. Subjects conditionally cooperate more often when free-riding imposes larger losses on the first-mover, or when free-riding provides smaller gains for oneself.

These findings provide two important implications. First, the within-subject variation of conditional cooperation with payoffs suggests that conditional cooperation should be viewed as an endogenous behavior arising from interaction between underlying motives and payoff variations, rather than a preference itself (Arifovic and Ledyard, 2012). A majority of subjects change their second-mover strategy when material payoffs change, and so classifications of individuals as conditional cooperators or free-riders should not be generalized to other games with different material payoffs.

Second, these results suggest that conditional cooperation may reflect underlying social preferences. The finding that strategies are sensitive to the cost imposed on the opponent as well as the gain to self suggests that a substantial proportion of subjects care not only their own material payoffs but also the other's material payoffs. Moreover, the way conditional cooperation varies with damage and gain is consistent with the predictions of several distributional preference models (e.g., Andreoni and Miller, 2002; Fehr and Schmidt, 1999).

To our knowledge, our study is the first study that empirically examines the within-subject variability of conditional cooperation when payoffs vary. Several questions follow naturally from our findings. One of these is whether individual distributional preferences determine the second-mover strategy in social dilemmas. In further research, we plan to employ a within-subject experiment consisting of two parts, where in the first-part individual distributional preferences are elicited from modified dictator games, and where in the second-part second-mover strategies for playing a sequential prisoner's dilemma are elicited. This would allow us to examine the hypothesis of whether conditional cooperation or free-riding is derived from individual distributional preferences.

References

Ahn, T., E. Ostrom, D. Schmidt, R. Shupp, and J. Walker. "Cooperation in PD games. Fear, Greed, and History of Play." *Public Choice*, 106, 2001, 137–155.

Andreoni, J., and J. Miller. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, 70(2), 2002, 737–753.

Arechar, A., S. Gächter, and L. Molleman. "Conducting interactive experiments online." *Experimental Economics*, 21, 2018, 99–131.

Arifovic, J., and J. Ledyard. "Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism." *Journal of Public Economics*, 96, 2012, 808–823.

Blanco, M., D. Engelmann, A.K. Koch, and H. T. Normann. "Preferences and beliefs in a sequential social dilemma: a within-subjects analysis."' *Games and Economic Behavior*, 87, 2014, 122–135.

Bolton, G.E., and A. Ockenfels. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1), 2000, 166–193.

Brandts, J., and G. Charness. "Hot vs. cold: sequential responses in simple experimental games." *Experimental Economics*, 2, 2000, 227–238.

Brandts, J., and G. Charness. "The strategy versus the direct-response method: a first survey of experimental comparisons." *Experimental Economics*, 14, 2011, 375–298.

Brandts, J., and A. Schram. "Cooperation and noise in public goods experiments: applying the contribution function approach." *Journal of Public Economics*, 79, 2001, 399–427.

Brosig, J., T. Riechmann, and J. Weimann. "Selfish in the end? An investigation of consistency and stability of individual behavior." MPRA Paper 2035, University Library of Munich, 2007.

Charness, G., and M. Rabin. "Understanding Social Preferences with Simple Tests." *The Quarterly Journal of Economics*, 117(3), 2002, 817–869.

Chaudhuri, A. "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature." *Experimental Economics*, 14, 2011, 47–83.

Clark, K., and M. Sefton. "The Sequential Prisoner's Dilemma: Evidence on Reciprocation." *The Economic Journal*, 111, 2001, 51–68.

Costa Lobo, M., and J. Curtice. "Personality Politics?: The Role of Leader Evaluations in Democratic Elections." Oxford University Press, 2015.

Cox, J.C., D. Friedman, and S. Gjerstad. "A Tractable Model of Reciprocity and Fairness." *Games and Economic Behavior*, 59(1), 2007, 17-45.

Eichenseer, M., and J. Moser. "Conditional cooperation: Type stability across games." *Economics Letters*, 188, forthcoming, 2020.

Fallucchi, F., R. A. Luccasen, and T. L. Turocy. "Identifying discrete behavioural types: a re-analysis of public goods game contributions by hierarchical clustering." *Journal of the Economic Science Association*, 5, 2019, 238–254.

Fehr, E., and K. M. Schmidt. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114, 1999, 817–868.

Fehr, E., and U. Fischbacher. "Social norms and human cooperation." *Trends in cognitive sciences*, 8(4), 2004, 185–190.

Fischbacher, U., S. Gächter and E. Fehr. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, 71(3), 2001, 397–404.

Gächter, S. "Conditional cooperation. Behavioral regularities from the lab and the field and their policy implications," In B. S. Frey & A. Stutzer (Eds.), Economics and psychology. A promising new cross-disciplinary field. Cambridge: MIT Press., 2007

Gächter, S., L. Gerhards, and D. Nosenzo. "The importance of peers for compliance with norms of fair sharing." *European Economic Review*, 97, 2017, 72–86.

Gächter, S., K. Lee, and M. Sefton. "Risk, Temptation, and Efficiency in Prisoner's Dilemmas." CeDEx Discussion Paper 2020-15, University of Nottingham, 2020.

Gächter, S., D. Nosenzo, E. Renner, and M. Sefton. "Who Makes a Good Leader? Social Preferences and Leading-by-Example." *Economic Inquiry*, 50(4), 2012, 953–967.

Giamattei, M., K.S. Yahosseini, S. Gachter, and L. Molleman. "LIONESS Lab: a free web-based platform for conducting interactive experiments online." *Journal of Economic Science Association*, forthcoming 2020.

Katuscak, P., and T. Miklánek. "What Drives Conditional Cooperation in Public Goods Games?." CERGE-EI Working Paper Series No.631, 2018.

Keser, C., and F. van Winden. "Conditional Cooperation and Voluntary Contributions to Public Goods." *The Scandinavian Journal of Economics*, 102(1), 2000, 23–39.

Kocher, M.G., T. Cherry, S. Kroll, R.J. Netzer, and M. Sutter. "Conditional cooperation on three continents." *Economics Letters*, 101(3), 2008, 175–178.

Kroh, M. "Measuring Left-Right Political Orientation: The Choice of Response Format." *Public Opinion Quarterly*, 71(2), 2007, 204–220.

Miettinen, T., M. Kosfeld, E. Fehr, and J.W. Weibull. "Revealed preferences in a sequential prisoners' dilemma: A horse-race between six utility functions." *Journal of Economic Behavior and Organization*, 173, 2020, 1-25.

Muller, L., M. Sefton, R. Steinberg, and L. Vesterlund. "Strategic behavior and learning in repeated voluntary contribution experiments."*Journal of Economic Behavior and Organization*, 67, 2008, 782–793.

Mullett, T., R. McDonald, and G. Brown. "Cooperation in public goods games predicts behavior in incentive-matched binary dilemmas: Evidence for stable prosociality."*Economic Inquiry*, 58(1), 2020, 67–85.

Schmidt, D., R. Shupp, J. Walker, T.K. Ahn, and E. Ostrom. "Dilemma games: game parameters and matching protocols."*Journal of Economic Behavior and Organization*, 46, 2001, 357–377.

Selten, R. Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes. In: Sauermann, H. (Ed.), Beiträge zur experimentellen Wirtschaftsforschung. J.C.B. Mohr (Paul Siebeck), Tübingen, 1967, 136–168.

Thöni, C., and S. Volk. "Conditional cooperation: Review and refinement." *Economics Letters*, 171, 2018, 37–40.

Vlaev, I., and N. Chater. "Game relativity: how context influences strategic decision making." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 2006, 131-149.

Volk, S., C. Thöni, and W. Ruigrok. "Temporal stability and psychological foundations of cooperation preferences." *Journal of Economic Behavior and Organization*, 81, 2012, 664–676.

Appendices

Appendix 3.A Theoretical Analysis

In sequential prisoner's dilemmas, two players sequentially decide whether to cooperate or defect. The first mover chooses to either cooperate or defect, and, after observing this choice, the second-mover then chooses to cooperate or defect. Players' earnings are maximized when both cooperate and each player receives R. However, if the first-mover cooperates, the second-mover maximizes her own earnings by defecting, in which case the first-mover receives S and the second-mover receives T. If the first-mover defects, the second-mover maximizes her own earnings by also defecting, so that each player receives P.

This Appendix provides the theoretical analysis to examine how the second-mover's choice (e.g., conditional cooperation, free-riding) is predicted and varies with payoffs. For theoretical analysis, we use two social preference models developed by Andreoni and Miller (2002) and Fehr and Schmidt (1999), which have been widely used to explain unselfish behavior across dictator games, prisoner's dilemma games, and public goods games.

First, Andreoni and Miller (2002) suggests the CES utility function given by

$$U_s = [\alpha \pi_s^{\rho} + (1-\alpha)\pi_o^{\rho}]^{1/\rho}$$

where a subject's utility depends on own earnings (π_s) and other's earnings (π_o). α indicates the relative weight on own earnings (the degree of fair-mindedness) and α ranges from 0 to 1, where higher α implies lower degree of fair-mindedness. The parameter ρ, which ranges from -∞ to 1, measures the curvature of indifference curves in (π_s, π_o) space. Indifference curves are L-shaped when ρ tends to -∞, so that utility maximisation will imply equalizing earnings. When $\rho = 1$, indifference curves are linear so that utility maximisation will imply allocating all income to self or all income to other (depending on the relative price of giving).

The CES model predicts that i) the second-mover cooperates in response to first-mover's cooperation if and only if the utility of mutual cooperation outweighs that of defecting against cooperator (i.e. $R^\rho > [\alpha T^\rho + (1-\alpha)S^\rho]$), and ii) the second-mover cooperates in response to first-mover's defection if and only if the utility of cooperating against defector outweighs that of mutual defection (i.e. $[\alpha S^\rho + (1-\alpha)T^\rho] > P^\rho$).

Therefore, the second-mover's utility maximizing strategy depends on preferences (α, ρ) and payoffs (T, R, P, S) as follows.

$$\begin{cases} \text{Always defect if } \alpha > max\{\frac{R^\rho - S^\rho}{T^\rho - S^\rho}, \frac{T^\rho - P^\rho}{T^\rho - S^\rho}\} \\ \text{Conditionally cooperate if } \frac{T^\rho - P^\rho}{T^\rho - S^\rho} < \alpha < \frac{R^\rho - S^\rho}{T^\rho - S^\rho} \\ \text{Unconditionally cooperate if } \alpha < min\{\frac{R^\rho - S^\rho}{T^\rho - S^\rho}, \frac{T^\rho - P^\rho}{T^\rho - S^\rho}\} \\ \text{Mismatch if } \frac{R^\rho - S^\rho}{T^\rho - S^\rho} < \alpha < \frac{T^\rho - P^\rho}{T^\rho - S^\rho} \end{cases}$$

Second, Fehr and Schmidt (1999, henceforth FS) provide a model based on the assumption that players have preferences that can be represented by utility functions of the following form.

$$U_s = \begin{cases} \pi_s - \beta \max\{\pi_s - \pi_o, 0\} & \text{if } \pi_s > \pi_o \\ \pi_s - \gamma \max\{\pi_o - \pi_s, 0\} & \text{if } \pi_s < \pi_o \end{cases}$$

, where $\beta \leq \gamma$ and $0 \leq \beta < 1$. β indicates the distaste of advantageous inequality, and γ indicates the distaste of disadvantageous inequality. When $\gamma = \beta = 0$, the utility is only affected by one's own payoff and does not have any relevance to other's payoff: therefore, such subject can be classified as purely self-interested. In case of $\gamma \geq \beta > 0$, a subject exhibits difference aversion in that he/she put more weights on his/her own payoffs when his/her payoff is smaller than other's and he put the positive weight on other's payoff when his/her payoff is greater than other's.

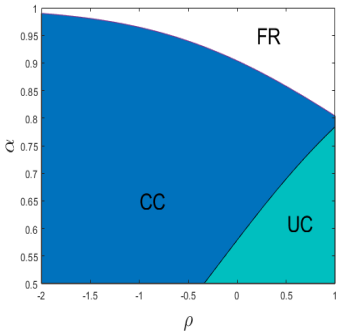
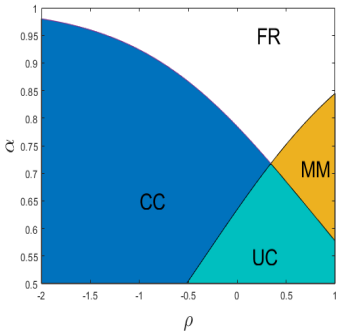
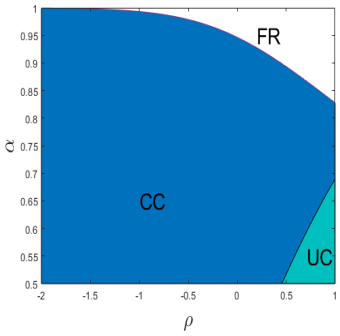
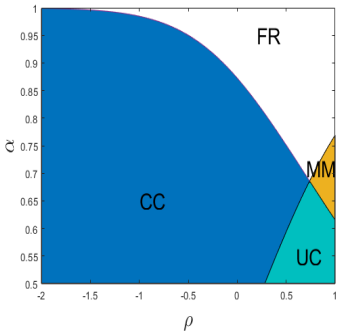
Given for payoffs, the second-mover's choices in sequential prisoner's dilemmas are determined by β as follows. Note that unconditional cooperation and mismatching are not predicted in the FS model because of the assumption of non-negativity in γ.

$$\begin{cases} \text{Always defect if } \beta < \frac{T-R}{T-S} \\ \text{Conditionally cooperate if } \beta > \frac{T-R}{T-S} \end{cases}$$

Both models predict that the likelihood of conditional cooperation increases with R but decreases with T and S. In terms of the indices of DAMAGE and GAIN, the likelihood of conditional cooperation increases with DAMAGE ($\frac{R-S}{R}$) but decreases with GAIN ($\frac{T-R}{R}$). Figures 3.A1-A2 illustrate the predicted second-mover's strategies of the CES and the FS models, respectively, in our eight sequential games. As shown in Figures, the likelihood of conditional cooperation increases with DAMAGE but decreases with GAIN. Reversely, the likelihood of free-riding decreases with DAMAGE but increases with GAIN.

Figure 3.A1: Predicted Second-mover's Strategies in the CES model

Panel (a). *high* EFF

	low GAIN	*high* GAIN
low DAMAGE		
high DAMAGE		

Panel (b). *low* EFF

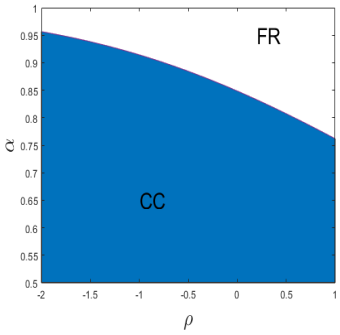
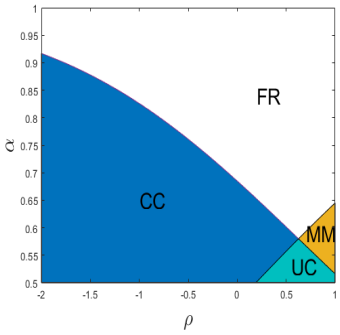
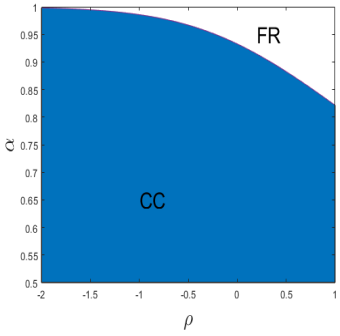
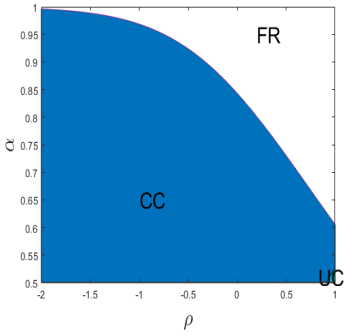
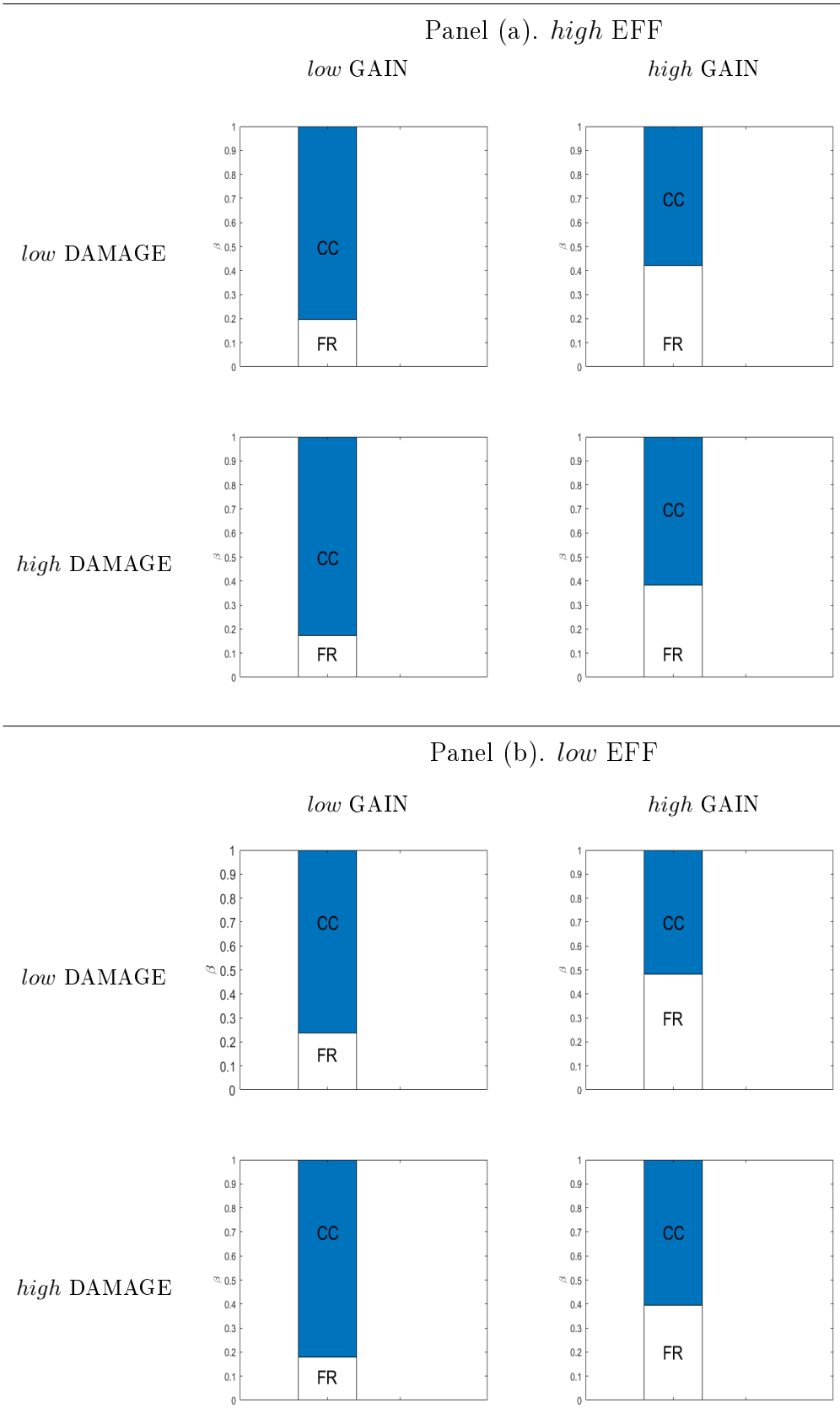
	low GAIN	*high* GAIN
low DAMAGE		
high DAMAGE		

Figure 3.A2: Predicted Second-mover's Strategies in the FS model



Appendix 3.B Additional Results

3.B.1 Second-mover's Decision

Tables 3.B1-B2 provide the full regression results for the determinants of conditional cooperation and free-riding, respectively.

Table 3.B1: Determinants of Conditional Cooperation

	(1) All	(2) Switcher	(3) All	(4) Switcher
DAMAGE	0.274**	0.356**	0.276**	0.374**
	(0.113)	(0.147)	(0.113)	(0.150)
GAIN	-0.245***	-0.323***	-0.263***	-0.348***
	(0.067)	(0.087)	(0.067)	(0.087)
EFF	-0.101	-0.133	-0.096	-0.127
	(0.073)	(0.096)	(0.076)	(0.101)
Round	-0.016**	-0.020**	-0.017**	-0.022**
	(0.007)	(0.009)	(0.007)	(0.009)
B = Coop.			-0.011	-0.013
			(0.027)	(0.036)
SecondThenFirst			-0.044	-0.046
			(0.031)	(0.041)
age			0.003	0.004
			(0.003)	(0.003)
female			-0.004	0.027
			(0.066)	(0.069)
politics			-0.004	-0.004
			(0.010)	(0.009)
Mturk exp.			0.071	0.094
			(0.063)	(0.059)
full-time			-0.052	0.068
			(0.063)	(0.062)

	(1) All	(2) Switcher	(3) All	(4) Switcher
higher			-0.071	0.013
education			(0.070)	(0.071)
income			-0.040	-0.015
$\geq$ \$50,000			(0.065)	(0.063)
White			0.130	0.021
			(0.082)	(0.078)
		Session - baseline: Session 1		
Session 2			0.161*	0.099
			(0.083)	(0.084)
Session 3			-0.104	-0.057
			(0.080)	(0.086)
Session 4			-0.072	-0.022
			(0.085)	(0.081)
Session 5			0.096	0.147
			(0.101)	(0.095)
Controls	*No*	*No*	*Yes*	*Yes*
Observations	848	624	808	600
Log-likelihood	-731.4	-656.0	-676.8	-610.6
BIC	1,604.4	1,447.2	1775.5	1624.2

Notes: Average marginal effects from panel multinomial logit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$. B=Coop. is an indicator variable for whether the cooperative action was labelled as option B. SecondThenFirst indicates the second-mover decision task is placed in the upper section and the first-mover decision task in the lower section of the screen. Mturk.exp indicates the subject has taken part in previous MTurk experiments more than "Once or Twice". Education (Less than a high school degree, High school Diploma, Vocational training, Attended College, Undergraduate, Graduate school, Doctoral or further), Ethnicity (American Indian or Alaska Native, Asian, Black of American African, Native Hawaiian or Other Pacific Islander, White, Latin American, Others), Employment (Full-time worker, Part-time worker, Unemployed, Student, Retired, Self-employed, Homemaker, Unable to work), Income (Less than \$29,999, \$30,000 $\sim$ \$49,999, \$50,000 $\sim$ \$69,999, \$70,000 $\sim$ \$89,999, \$90,000 or more) Politics (0 to 10, 0 means the most left and 10 means the most right). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 3.B2: Determinants of Free Riding

	(1) All	(2) Switcher	(3) All	(4) Switcher
DAMAGE	-0.260**	-0.337**	-0.227*	-0.308*
	(0.118)	(0.157)	(0.118)	(0.160)
GAIN	0.155**	0.200**	0.160**	0.211**
	(0.069)	(0.092)	(0.068)	(0.093)
EFF	0.008	0.007	-0.012	-0.017
	(0.070)	(0.093)	(0.070)	(0.097)
Round	0.021***	0.028***	0.021***	0.029***
	(0.006)	(0.008)	(0.006)	(0.008)
B = Coop.			-0.011	-0.015
			(0.032)	(0.044)
SecondThenFirst			0.043	0.050
			(0.030)	(0.041)
age			-0.004	-0.006*
			(0.003)	(0.004)
female			-0.007	0.010
			(0.062)	(0.062)
politics			-0.009	-0.011
			(0.011)	(0.013)
Mturk exp.			-0.044	-0.043
			(0.065)	(0.069)
full-time			0.077	-0.024
			(0.062)	(0.067)
higher			0.094	0.043
education			(0.068)	(0.072)
income			0.056	-0.003
$\geq$ \$50,000			(0.064)	(0.069)

	(1) All	(2) Switcher	(3) All	(4) Switcher
White			-0.159**	-0.002
			(0.077)	(0.075)
		Session - baseline: Session 1		
Session 2			-0.184**	-0.133
			(0.089)	(0.103)
Session 3			0.126	0.063
			(0.083)	(0.094)
Session 4			0.135	0.096
			(0.089)	(0.095)
Session 5			0.020	-0.004
			(0.106)	(0.101)
Controls	*No*	*No*	*Yes*	*Yes*
Observations	848	624	808	600
Log-likelihood	-731.4	-656.0	-676.8	-610.6
BIC	1,604.4	1,447.2	1775.5	1624.2

Notes: Average marginal effects from panel multinomial logit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$. B=Coop. is an indicator variable for whether the cooperative action was labelled as option B. SecondThenFirst indicates the second-mover decision task is placed in the upper section and the first-mover decision task in the lower section of the screen. Mturk.exp indicates the subject has taken part in previous MTurk experiments more than "Once or Twice". Education (Less than a high school degree, High school Diploma, Vocational training, Attended College, Undergraduate, Graduate school, Doctoral or further), Ethnicity (American Indian or Alaska Native, Asian, Black of American African, Native Hawaiian or Other Pacific Islander, White, Latin American, Others), Employment (Full-time worker, Part-time worker, Unemployed, Student, Retired, Self-employed, Homemaker, Unable to work), Income (Less than \$29,999, \$30,000 $\sim$ \$49,999, \$50,000 $\sim$ \$69,999, \$70,000 $\sim$ \$89,999, \$90,000 or more) Politics (0 to 10, 0 means the most left and 10 means the most right). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

In Table 3.B3, we report the regression results that employ the same econometric specifications of Table 3.3, but use the indices of RISK and TEMPT. The overall patterns of CC (Panel A) and FR (Panel B) with RISK and TEMPT are similar to the results using DAMAGE and GAIN.

Table 3.B3: Determinants of CC (Panel A) and FR (Panel B)

	A. Conditional Cooperation (CC)			
	(1) All	(2) Switcher	(3) All	(4) Switcher
RISK	0.186**	0.243**	0.189***	0.257***
	(0.073)	(0.096)	(0.073)	(0.097)
TEMPT	-0.471***	-0.621***	-0.504***	-0.667***
	(0.129)	(0.167)	(0.129)	(0.167)
EFF	-0.027	-0.037	-0.022	-0.028
	(0.063)	(0.083)	(0.066)	(0.089)
Round	-0.016**	-0.020**	-0.017**	-0.022**
	(0.007)	(0.009)	(0.007)	(0.009)
	B. Free Riding (FR)			
	(1) All	(2) Switcher	(3) All	(4) Switcher
RISK	-0.166**	-0.215**	-0.142*	-0.194*
	(0.078)	(0.104)	(0.078)	(0.105)
TEMPT	0.299**	0.385**	0.306**	0.406**
	(0.132)	(0.177)	(0.131)	(0.179)
EFF	-0.063	-0.084	-0.073	-0.099
	(0.061)	(0.080)	(0.061)	(0.084)
Round	0.021***	0.028***	0.021***	0.029***
	(0.006)	(0.008)	(0.006)	(0.008)
Controls	*No*	*No*	*Yes*	*Yes*
Observations	848	624	808	600
Log-likelihood	-730.4	-655.0	-675.8	-609.5
BIC	1,602.5	1,445.2	1773.4	1622.0

Note: Average marginal effects from panel multinomial logit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$. Controls: the variables of demographic information and session effects.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

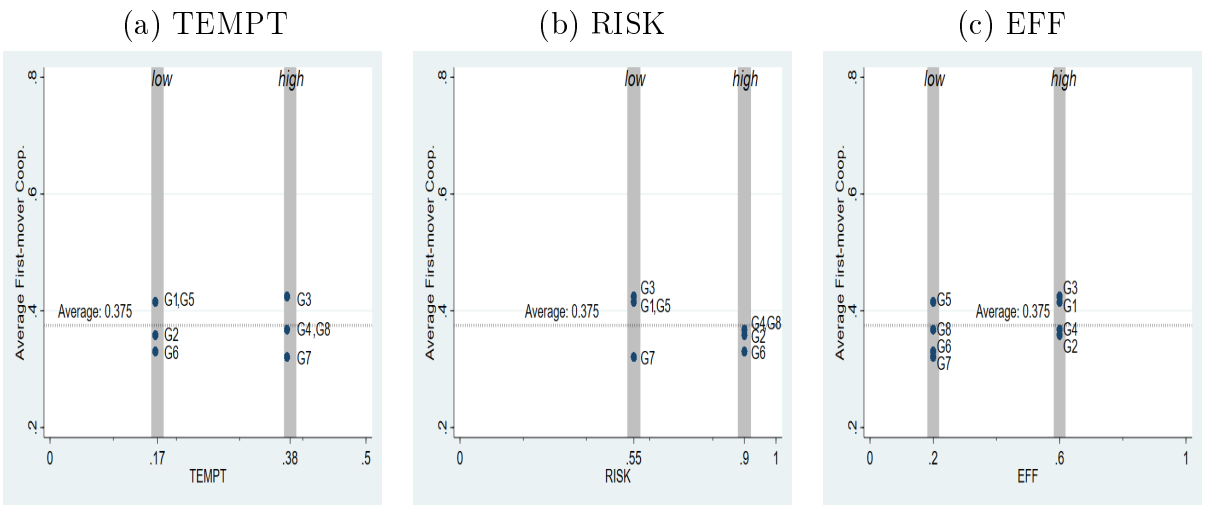
When RISK increases, subjects are more likely to conditionally cooperate but less likely to free-ride. This suggests that the opponent's earnings are the influential determinants of whether to conditionally cooperate or to free-ride. TEMPT has the opposite effect to DAMAGE: when TEMPT increases, subjects are less likely to conditionally cooperate but more likely to free-ride. This indicates that the willingness to reciprocate to the first-mover's cooperation becomes weaker with the gains from free-riding.

3.B.2 First-mover's Decision

In this section, we focus on how first-mover's cooperation is affected by payoffs and subjects' own decision as a second-mover. In the sequential game, the first-mover's decision depends on not only one's own subjective beliefs about the second-mover's cooperativeness but also the first-mover's individual preferences.

Figure 3.B1 illustrates aggregate first-mover's cooperation rates depending on the level of each payoff index. First of all, first-mover's cooperation exhibits an unclear relationship with TEMPT and EFF. We do not find significant differences in first-mover's cooperation rates between *low* and *high* levels of TEMPT (Figure B1(a)) and EFF (Figure B1(c)) indices. Second, first-mover's cooperation tend to decrease with RISK (Figure B1(b)). Except for one game (Game 7), average frequency of first-mover's cooperation decreases when RISK increases.

Figure 3.B1: Aggregate First-mover's Cooperation Rates with TEMPT, RISK, and EFF

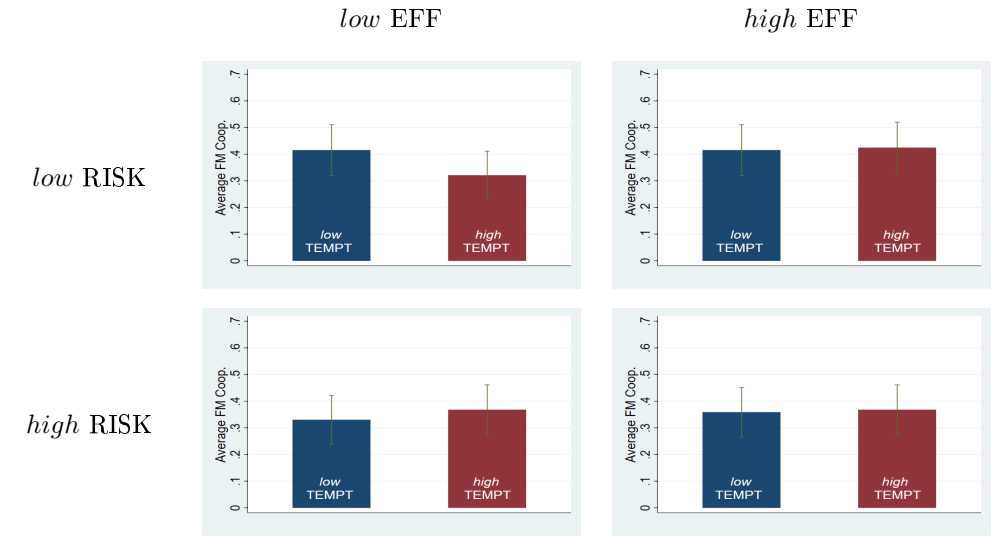


Note: TEMPT: $\frac{T-R}{T}$, RISK: $\frac{P-S}{P}$, EFF: $\frac{R-P}{R}$.

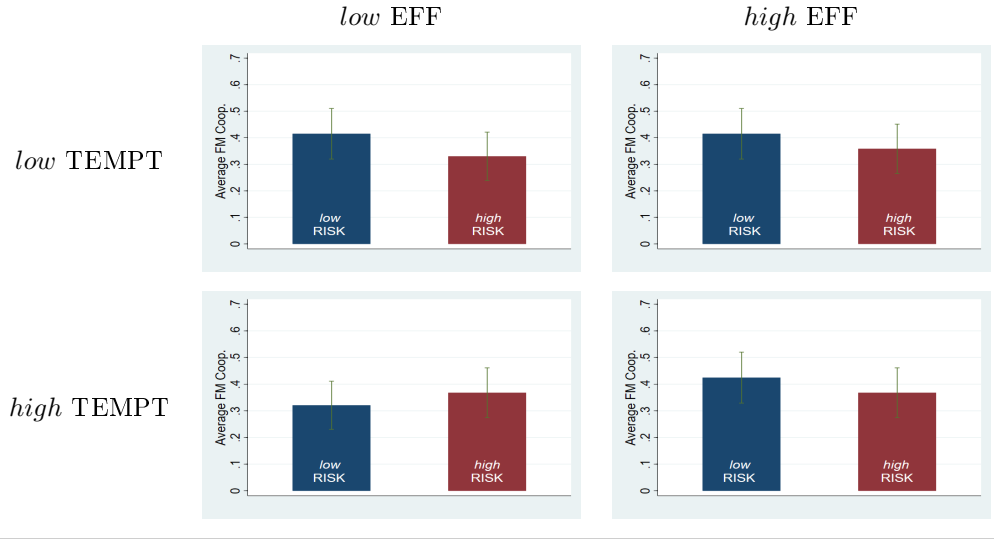
Figure 3.B2 illustrates how first-mover's cooperation rates vary with the variation in each payoff index holding other payoff indices constant. First, we find that average first-mover's cooperation rates marginally vary with the payoff indices. In Panel (a), we see that first-mover's cooperation rates for the *low* TEMPT are on average only 1 percentage point higher than for the *high* TEMPT (with the absolute differences ranging from 1 percentage points to 9 percentage points). First-mover's cooperation rates for the *low* RISK are on average 4 percentage points higher than for the *high* RISK (with the absolute differences ranging from 5 percentage points to 8 percentage points). First-mover's cooperation rates for the *low* EFF are on average 3 percentage points lower than for the *high* EFF (with the absolute differences ranging from 0 percentage points to 10 percentage points).

Figure 3.B2: First-mover's Cooperation Rates between PDs

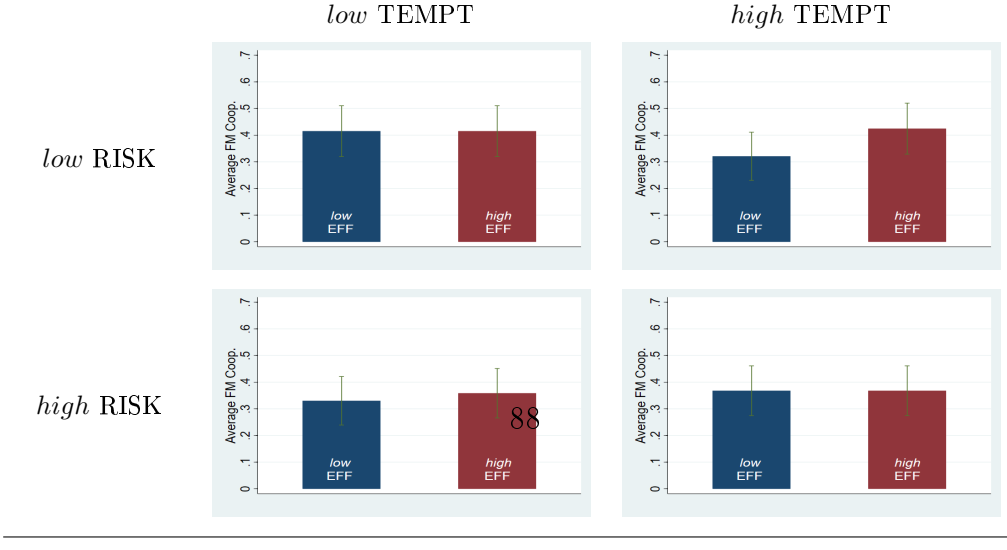
Panel (a). *low* vs *high* TEMPT



Panel (b). *low* vs *high* RISK



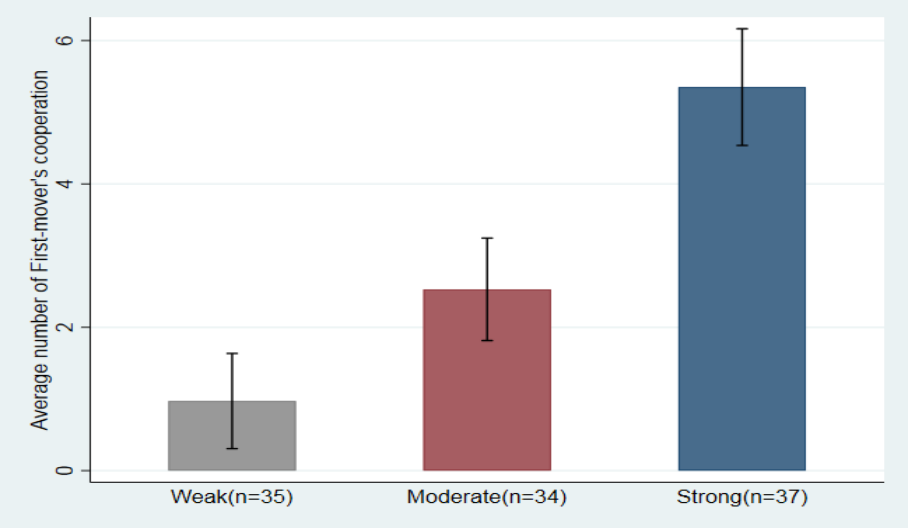
Panel (c). *low* vs *high* EFF



Next, we measure the degree of cooperativeness by counting the number of positive reciprocation (i.e. cooperates if the first-mover cooperates) across eight games. We classify subjects into three groups - Strong Reciprocator, Moderate Reciprocator, and Weak Reciprocator - to examine how the degree of cooperativeness affects the first-mover's decision. Subjects who reciprocate to first-mover's cooperation at least 6 times out of 8 games are classified as Strong Reciprocator (henceforth SR), and 35% of subjects fall into this category. 32% of subjects are classified as Moderate Reciprocator (henceforth MR) if subjects reciprocate to first-mover's cooperation more than twice but less than 6 times. Lastly, subjects who reciprocate to first-mover's cooperation at most twice are classified as Weak Reciprocator (henceforth WR), and 33% of subjects were included in this category.

First, the degree of cooperativeness is strongly associated with the frequency of first-mover's cooperative choice. Figure 3.B3 illustrates the average number of first-mover's cooperative choice across SR, MR, and WR subjects. SR subjects cooperate as a first-mover on average 5.3 games, while MR and WR subjects cooperate as a first-mover only 2.5 and 1 games, respectively. We can strongly reject the hypothesis that the frequency of first-mover's cooperation does not differ between SR, MR, and WR subjects (Kruskal-Wallis test: χ^2=47.14, $p < 0.001$). We also find that the frequency of first-mover's cooperation between each category and another category - SR and MR, MR and WR, SR and WR - does significantly differ using Mann-Whitney U test ($p = 0.001$).

Figure 3.B3: Average Number of First-mover's Cooperation



Note: Lines represent 95% confidence intervals.

Second, CC and FR reveal the asymmetric effect on first-mover's choice. In only 18% of games, subjects who free-ride cooperate as a first-mover. Therefore, a vast majority of subjects who are classified as free-riders defects as a first-mover. Contrary to this, subjects who are classified as conditional cooperators cooperate as a first-mover in more than half of games (56%). Still, subjects who exhibit conditional cooperativeness do not cooperate as a first-mover in the remaining 44% of games. This implies that strategic uncertainty deters conditional cooperators to cooperate as a first-mover.

Regression analysis

Table 3.B4 provides the regression results that examine the determinants of first-mover's cooperation. We find that second-mover strategy and degree of cooperativeness are linked to first-mover's choice. Column (1) only include the variables of payoff indices. Columns (2)-(3) include the second-mover strategy in addition to the regressors in column (1). Columns (4)-(5) include the variables related to the degree of cooperativeness in addition to the regressors in

column (1).

First of all, we cannot find any significant relation between payoff indices and first-mover's choice in all columns. Although subjects tend to less likely to cooperate as a first-mover as RISK increases, TEMPT increases, or EFF decreases, the effect sizes are marginal and insignificant. Second, we find the significant association between second-mover strategy and first-mover's cooperation. Conditional cooperators are more likely to cooperate as a first-mover by 8.1 percentage points than subjects who choose other strategies (column (2)). In contrast, free-riders are less likely to cooperate as a first-mover by 13.5 percentage points than subjects who choose non-FR strategies (column (3)).

Columns (4)-(5) clearly show that the degree of cooperativeness significantly affects first-mover's choice. Both the degrees of CC (the frequency of CC in eight games) and FR (the frequency of FR in eight games) exhibit large marginal effect sizes - 0.498 and -0.589 - , and are significant at the 1% level. The marginal effects can be interpreted as follows: i) the probability of first-mover's cooperation who always conditionally cooperate are higher by 49.8 percentage points than subjects who never conditionally cooperates, and ii) the probability of first-mover's cooperation who always free-ride are lower by 58.9 percentage points than subjects who never free-ride.

Overall, our study supports the previous literature that finds the relation between second-mover's decision and first-mover's decision (Brandts and Charness, 2000; Blanco, Engelmann, Koch and Normann, 2014; Gächter, Nosenzo, Renner and Sefton, 2012). In line with these studies, we find that second-mover strategies and the degree of cooperativeness are significantly associated with first-mover's cooperation.

Table 3.B4: Determinants of First-mover's Cooperative Choice

	(1)	(2)	(3)	(4)	(5)
RISK	-0.079	-0.096	-0.097	-0.079	-0.079
	(0.061)	(0.064)	(0.065)	(0.060)	(0.060)
TEMPT	-0.021	0.024	0.024	-0.018	-0.014
	(0.110)	(0.116)	(0.120)	(0.109)	(0.108)
EFF	0.093	0.097	0.088	0.090	0.088
	(0.062)	(0.065)	(0.066)	(0.061)	(0.061)
1 if CC		0.081**			
		(0.036)			
1 if FR			-0.135***		
			(0.045)		
Frequency of CC				0.498***	
				(0.064)	
Frequency of FR					-0.589***
					(0.053)
Round	-0.026***	-0.026***	-0.025***	-0.026***	-0.026***
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
Controls	*No*	*No*	*No*	*No*	*No*
Observations	848	848	848	848	848
Log-likelihood	-411.6	-408.8	-404.1	-394.0	-383.6
BIC	863.6	864.7	855.3	835.1	814.3
Pseudo R^2	0.035	0.042	0.053	0.076	0.101

Note: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$. Controls: the variables of demographic information, task characteristics, and session effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Appendix 3.C Experimental Instructions

Welcome

Thank you for accepting this HIT. To complete this HIT, you must make some decisions. Including the time for reading these instructions, the HIT will take about 30 minutes to complete. If you are using a desktop or laptop to complete this HIT, we recommend that you maximize your browser screen (press F11) before you start.

It is important that you complete this HIT without interruptions. During the HIT, please **do not close this window or get distracted from the task.** If you close your browser or leave the task, you will not be able to re-enter and we will not be able to pay you.

In this HIT, you will be matched with one other participant. Each of you will make decisions for 8 decision situations. In each situation, each of you will earn Tokens depending on your decisions.

At the end of the HIT, one of the decision situations will be randomly chosen. Your earnings from this situation will be converted from Tokens to Dollars at a rate of 1**00 Tokens = $ 1.** This will be added to **your participation fee of $1.00.** Depending on your decisions, you may make up to $8.00 more in addition to the $1.00 participation fee. In the same way, Tokens earned by the person matched with you in that same situation will also be converted to Dollars at a rate of 100 Tokens = $ 1.

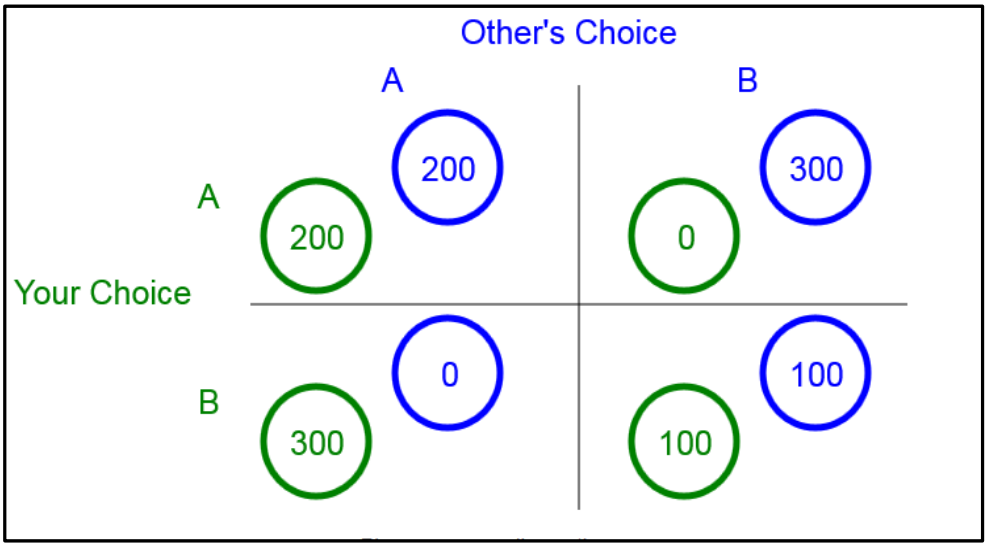
You will receive a code to collect your payment via MTurk upon completion.

Please click "Continue" to start the HIT.

INSTRUCTIONS

The HIT consists of 8 decision situations.

Each decision situation will be presented on a screen like the **example screen** below.



You and the other person will be making choices between **A** and **B**. Your earnings are the values in the green circle, and the other person's earnings are the values in the blue circle. The table is read as follows:

- If you choose A and the other person chooses A, you will earn 200 Tokens and the other person will earn 200 Tokens.
- If you choose A and the other person chooses B, you will earn 0 Tokens and the other person will earn 300 Tokens.
- If you choose B and the other person chooses A, you will earn 300 Tokens and the other person will earn 0 Tokens.
- If you choose B and the other person chooses B, you will earn 100 Tokens and the other person will earn 100 Tokens.

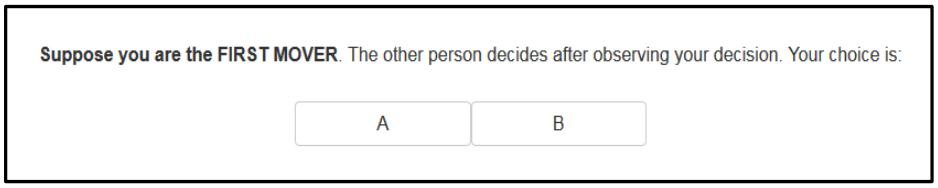
Please note that the values in the table will differ in each decision situation.

Tasks

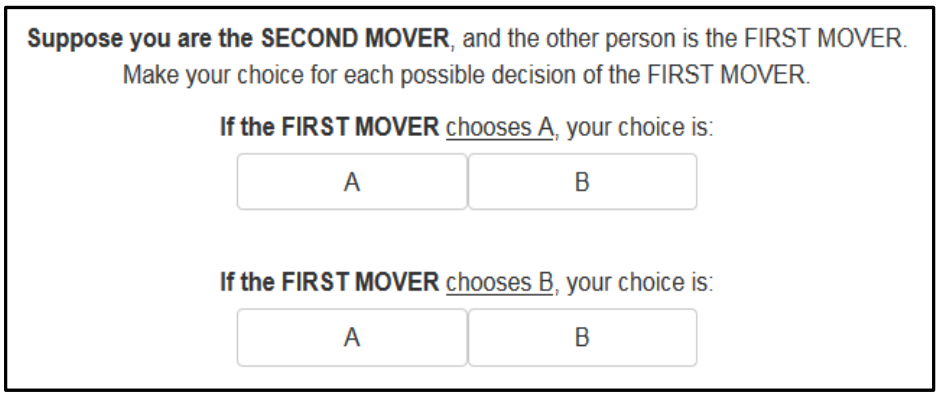
In each decision situation, you must complete **two types** of tasks, which we will refer to below as the "FIRST MOVER's decision" and "SECOND MOVER's decision". The FIRST MOVER decides first whether to choose A or B. The SECOND MOVER is then informed of the FIRST MOVER's decision. The SECOND MOVER then decides whether to choose A or B.

We want to know what you would do in the role of the FIRST MOVER and what would you do in the role of the SECOND MOVER. Thus you will be prompted to make decisions in both roles.

● For the "FIRST MOVER's decision" task, you will see the following screen and you must choose A or B:



● For the "SECOND MOVER's decision" task, You will see the following screen and you must choose A or B in two possible cases: (1) if the FIRST MOVER chooses A (2) if the FIRST MOVER chooses B



During the HIT, you will not receive any feedback on the other person's choice or the outcomes of the decision situations.

Your dollar earnings

On completion of the HIT, you will be paid your participation fee of $ 1.

In addition, one of the decision situations will be randomly chosen for your additional dollar earnings. The computer will randomly choose either you or the other person to be the first-mover. If you are chosen to be the first-mover, your first-mover's decision will be matched with the second-mover's decision of the other person. If the other person is chosen to be the first-mover, your second-mover's decision will be matched with the first-mover's decision of the other person. Your earnings and the other person's earnings will be determined depending on choices of you and the other person in that situation. Two examples should make this clear.

Example 1. Assume that **the computer randomly selects you to be the first-mover. This implies that your payoff relevant decision will be your first-mover's decision.** Assume that you choose A as the first-mover's decision in the above example screen. Assume that the other person matched with you makes the following second-mover's decisions: he/she chooses A if you choose A, and chooses B if you choose B. As a consequence, you will earn 200 Tokens and the other person will earn 200 Tokens.

Example 2. Assume that **the computer randomly selects the other person to be the first-mover. This implies that your payoff relevant decision will be your second-mover's decision.** Assume that you make the following second-mover's decisions: you choose B if the FIRST MOVER chooses A, and choose B if the FIRST MOVER chooses B in the above example screen. Assume that the other person matched with you chooses A as the first-mover's decision. As a consequence, you will earn 300 Tokens and the other person will earn 0 Tokens.

At the end of the HIT

On completion of the HIT, one of the decision situations will be randomly chosen as explained above. You will be informed of your choices and earnings for that decision situation, and you will be paid these earnings in addition to your participation fee.

Note that we will not be able to pay you if you do not complete the HIT. If the person you are matched with does not complete the HIT, the computer will randomly select one of the four possible earnings in the randomly chosen decision situation, and you will be paid these earnings in addition to your participation fee.

Your participation fee and the additional earnings will be paid to you within two working days.

Chapter 4

Distributional Preferences and Conditional Cooperation in Sequential Prisoner's Dilemmas

4.1 Introduction

Conditional cooperation is a widely observed behavioral pattern in social dilemmas. Many people cooperate in response to others' cooperation even though non-cooperation maximizes own earnings. This pattern of behavior is clear in numerous social dilemma experiments. In an experiment examining voluntary contributions to a public good, where subjects were able to make contributions conditional on others' contributions, Fischbacher, Gächter and Fehr (2001) found that many subjects match the contributions of others. Similar patterns of contributing if and only if others do so also have been found in repeated public good games (e.g., Fischbacher and Gächter, 2010; Gächter, Kölle and Quercia, 2017; Keser and van Winden, 2000). In sequential prisoner's dilemmas, where a second-mover maximizes earnings by defecting regardless of the first-mover's decision, Clark and Sefton (2001) find that many second-movers cooperate in response to first-mover cooperation.

A widely used approach to explain unselfish behavior in various contexts, including conditional cooperation in social dilemmas, uses theories of social preferences. Models of social preferences (e.g., Andreoni and Miller, 2002; Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Cox et al., 2007; Fehr and Schmidt, 1999) allow preferences to depend on others' earnings as well as own earnings, and incorporate motives such as desires to help others, to equalize earnings, or maximize combined earnings. In naturally occurring situations, either one or all of these motives could explain why people donate to the needy, or engage in pro-social acts.[1]

In this study, we assume social preferences take the form of constant elasticity of substitution (henceforth CES) utility function over own and others' earnings (Andreoni and Miller, 2002). This specification has two parameters which measure the degree of fair-mindedness and the concern for equality versus efficiency. We estimate the parameters of this social preference model, and examine how well these parameters predict conditional cooperation in sequential prisoner's dilemmas.

The rationale for employing Andreoni and Miller's model is that conditional cooperation in a standard payoff-symmetric social dilemma could be plausibly explained by either one or all of the following motives: fair-mindedness, the concern for equality, and the concern for efficiency. For example, individuals may conditionally cooperate because they do not want to exploit others by free-riding. Conditional cooperation could also arise from concerns for equality as conditional cooperation provides an egalitarian outcome. Similarly, con-

[1] Several experimental studies (Bellemare, Kröger and Van Soest, 2008; Beranek, Gächter and Cubitt, 2015; Bruhin, Fehr and Schunk, 2019; Cabrales, Miniaci, Piovesan and Ponti, 2010; Charness and Rabin, 2002; Fisman, Kariv and Markovits, 2007) empirically estimate parameters of various social preference models and find that a non-negligible proportion of subjects exhibit a taste for altrusim, equality, or efficiency.

cerns for efficiency may induce individuals to conditionally cooperate because cooperation provides the most efficient outcome.

To examine how well these motives explain conditional cooperation, we conduct an experiment consisting of two parts a week apart. In part 1 (week 1), we use the method developed by Andreoni and Miller (2002) and Fisman et al. (2007) to estimate distributional preferences: subjects make allocations between *oneself* and *other* in fifty rounds of modified dictator games, where continuous linear choice sets are provided and the relative price of giving (i.e. the relative price of other's earnings in terms of their own earnings) varies across games. We use these data to estimate the degree of fair-mindedness and equality-efficiency tradeoffs of a CES utility function.

In part 2 (week 2), subjects make decisions in twenty-four sequential prisoner's dilemmas where payoffs vary across games. We elicit second-mover choices using a strategy method so that we can observe how a second-mover responds to both cooperation and defection by a first-mover. Conditional cooperation is observed when the second-mover responds to defection with defection and responds to cooperation with cooperation. Using the individual-level distributional preference estimates from part 1, we generate predicted second-mover's strategies in each game for each individual. We then compare the predictions with the elicited strategies.

Our findings are three-fold. First, we show that distributional preferences help predict second-mover strategies in the sequential prisoner's dilemmas. Among 2,568 strategies (i.e. 107 subjects $\times$ 24 games), we correctly predict 55.8% of strategies. At the individual level, distributional preferences correctly predict at least 12 of 24 strategies for 56.1% of subjects. For comparison, we employ two other distributional preference models (Charness and

Rabin, 2002; Fehr and Schmidt, 1999), a random choice generator, and a self-interested model as benchmarks. The distributional preference models show similar performance in predicting second-mover strategy, whereas a random choice generator, which provides less than a 1% chance of correctly predicting 12 or more of an individual's 24 strategies, falls behind to distributional preference models. The self-interested model, which predicts that subjects are always free-rider regardless of payoffs, correctly predict at least 12 of 24 strategies for 64.5% of subjects. Although the distributional preference models do not outperform the self-interested model, we find that a non-negligible proportion of subjects' strategies are predicted by their distributional preferences.

Second, we find a clear positive relationship between fair-mindedness and conditional cooperation. Subjects who have a higher relative weight on other's earnings are more likely to conditionally cooperate. In contrast, equality-efficiency tradeoffs have an insignificant effect on decisions whether to conditionally cooperate or to free-ride. Both equality-oriented and efficiency-oriented subjects are more likely to conditionally cooperate than self-interested subjects, but we do not find significant differences between equality-oriented subjects and efficiency-oriented subjects in choosing conditional cooperation.

Third, we observe that fair-mindedness also induces subjects to cooperate as a first-mover. Subjects are more likely to cooperate as a first-mover when they put a higher relative weight on other's earnings. This suggests that distributional preferences play a role even under strategic uncertainty, although the effect size of concerns for others' earnings is substantially lower in first-movers' decisions than in second-movers' decisions.

Our study contributes to a growing body of evidence that individual decision-making is influenced by social preferences. For example, see Cooper and Kagel

(2016) for a review of experimental studies on trust games, bargaining games and gift exchange games. Our study differs from previous studies that examine how social preferences affect cooperation in the following two aspects. First, our experiment studies behavior across a broad set of payoff parameterizations in sequential prisoner's dilemmas. Most previous studies (e.g., Blanco et al., 2011; Hedegaard et al., 2019; Mullett et al., 2020; Teyssier, 2012) employed a particular payoff matrix, whereas we systematically vary payoffs across twenty-four sequential prisoner's dilemma games so that the likelihood of conditional cooperation varies in each game.

Second, we estimate the parameters of underlying distributional preference models which allow us to make precise predictions about whether subjects will conditionally cooperate in the prisoner's dilemma games. In contrast to several previous studies (e.g., Hedegaard et al., 2019; Mullett et al., 2020), which used one game to categorize subjects (e.g., as altruists, inequality-averse, selfish, free-riders, etc.) and then examined behavior in a related game, our approach allows us to test how well estimated distributional preference parameters predict behavior. This is important as we have shown in related work (see Chapter 3), that the category assigned to a given subject is sensitive to the payoff parameters used in the game.

The remainder of the chapter is organized as follows. In Section 4.2, we review findings on social preferences and conditional cooperation from the related literature. Section 4.3 describes the distributional preference model, how we estimate individual-level distributional preferences, and how we predict behavior in the sequential prisoner's dilemma. In Section 4.4, we explain our experimental design and procedures. Section 4.5 presents our results, and we conclude in Section 4.6.

4.2 Related Literature

We are aware of four previous laboratory experimental studies that examine the relationship between social preferences and conditional cooperation.[2] Two of these studies (Blanco et al., 2011; Teyssier, 2012) measured the parameters of social preference models from choices in ultimatum and modified dictator games and then examined how those preferences are associated with conditional cooperation. One study (Guan et al., 2016) measured the degree of inequality aversion from the amount of tokens used for purchasing a veto in the experiment, where a veto can change the unequal rule, then examine how the degree of inequality aversion is associated with conditional cooperation. The last study (Katuscak and Miklánek, 2018) did not directly measure social preferences, but employed several treatments in public goods games to examine the relationship between social preferences and conditional cooperation.

Blanco et al. (2011) conducted a within-subject experiment consisting of four games: an ultimatum game, a modified dictator game, a sequential prisoner's dilemma, and a public goods game. From the ultimatum game and modified dictator game, the authors elicited the Fehr and Schmidt (1999) model parameters of disadvantageous and advantageous inequality aversion, respectively. This study finds that the advantageous inequality aversion parameter helps predict the second-mover's cooperation in the sequential prisoner's dilemma. Teyssier (2012), which used the elicitation methods of Blanco et al. to measure inequality aversion, also reports that subjects who reveal

[2] Note that several studies examine the relationship between social preferences and cooperation in simultaneous social dilemmas: See Ghidoni et al. (2019), Hedegaard et al. (2019), Haesevoets et al. (2015), and Yamagishi et al. (2013) for a discussion. Miettenen et al. (2020) examines the relationship between social preferences, elicited from decisions as a second-mover in a sequential prisoner's dilemma, and first-mover cooperation. In our study, we focus on social preferences and *conditional cooperation* (i.e. how a second-mover responds to cooperation by a first-mover).

stronger advantageous inequality aversion are more likely to conditionally cooperate in the sequential public goods game than subjects who are less averse to advantageous inequality.

Guan et al. (2016) implemented a within-subject experiment consisting of two parts: i) a veto purchase experiment, which combines a dictator game and an ultimatum game, ii) a two-stage public goods game (Fischbacher and Gächter, 2010). The authors measure the degree of inequality aversion from the veto purchase experiment; and then examine how conditional cooperation is influenced by the degree of inequality aversion. They found that stronger inequality-averse subjects are more likely to conditionally cooperate in the public goods game, although the effect of inequality aversion is small and only marginally significant (i.e. at the 10% level).

Katuscak and Miklánek (2018) implemented a within-subject public goods experiment. This study aimed to separate the roles of inequality aversion, conformity and reciprocity for conditional cooperation. The authors reported that conditional cooperation is partially explained by inequality aversion and conformity, and hardly by reciprocity.

Overall, these studies suggest that various forms of social preference, especially concerns for equality, affect conditional cooperation. However, there are two limitations. First, most of these studies (Blanco et al., 2011; Guan et al., 2016; Teyssier, 2012) only used the model of inequality aversion. Therefore, these studies cannot examine how efficiency concerns influence conditional cooperation. Second, all of these studies examine how estimated parameters affect play in a given one-shot social dilemma game with fixed payoff parameters. Thus, the predictive power of distributional preferences for conditional cooperation is not tested under a broad set of payoff parameterizations.

To address these limitations, our study has the following important features. First, we use a CES utility function that can capture fair-mindedness and equality-efficiency tradeoffs. Therefore, our study allows us to explore how preferences weighted toward equality or efficiency affect conditional cooperation. Second, we examine the predictive power of estimated preferences across twenty-four sequential prisoner's dilemma games with varying payoffs.

4.3 Estimating Distributional Preferences and Predicting Conditional Cooperation

In this study, we estimate preferences by examining how a subject distributes budgets between him/herself and another subject. In doing this we mainly use a model of social preferences in which subjects maximize a CES utility function (see Andreoni and Miller, 2002; Fisman et al., 2007). We then use the estimated parameters of individual subjects to predict their second-mover strategy in sequential prisoner's dilemma games.[3]

4.3.1 Distributional Preferences

The CES utility function has the form

$$U_s = [\alpha \pi_s^{\rho} + (1-\alpha)\pi_o^{\rho}]^{1/\rho} \tag{4.1}$$

where a subject's utility depends on own earnings (π_s) and other's earnings (π_o). α indicates the relative weight on own earnings (the degree of fair-mindedness) and α ranges from 0 to 1, where higher α implies lower degree of

[3] Although this design is tailored to estimating the CES social preference model, we can also estimate the parameters of other distributional preference models from the data, and we report the results of this in Section 4.5.4.

fair-mindedness. The parameter ρ, which ranges from -∞ to 1, measures the curvature of indifference curves in (π_s, π_o) space. Indifference curves are L-shaped when ρ tends to -∞, so that utility maximisation will imply equalizing earnings. When $\rho = 1$, indifference curves are linear so that utility maximisation will imply allocating all income to self or all income to other (depending on the relative price of giving). In general, following Fisman et al. (2017), we classify distributional preferences as efficiency-oriented when $0 < \rho \leq 1$ and equality-oriented when $\rho < 0$.

4.3.2 Econometric Specification

We estimate individual-level parameters of the CES utility function from each subject's decisions in fifty rounds of modified dictator games. In the modified dictator game, the allocation task is presented graphically on a computer screen. A subject can decide of how much to allocate for him/herself (π_s) and how much to allocate to other (π_o), subject to a budget constraint $p_s\pi_s + p_o\pi_o = m$, where m denotes the endowment and p_o/p_s is the relative price of giving. In each modified dictator game, the relative price of giving and income are randomly determined by the computer at the individual level.

From the CES utility function (equation (4.1)), the CES demand function is given by

$$\pi_s(p_s, p_o, m) = \frac{g}{p^r + g}\frac{m}{p_s}$$

where $r = -\rho/(1-\rho)$, $g = [\alpha/(1-\alpha)]^{1/(1-\rho)}$, and $p = p_o/p_s$. This generates the following econometric specification for each subject i:

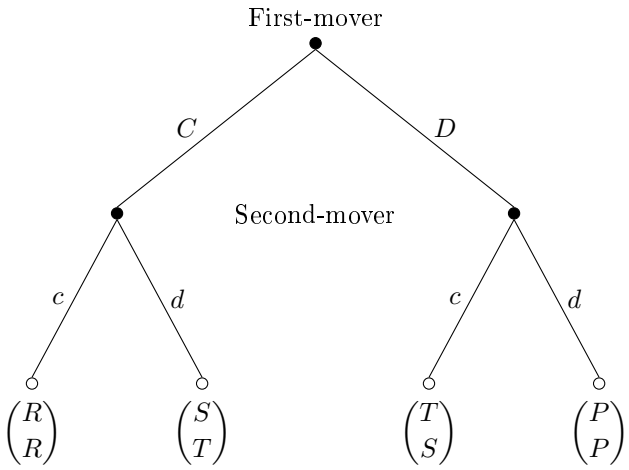
$$\frac{p_{s,i}^t \pi_{s,i}^t}{m_i^t} = \frac{g_i}{p_i^{t\,r_i} + g_i} + \epsilon_i^t$$

where $t \in \{1, ..., 50\}$ denotes the round of the modified dictator games and ϵ_i^t is assumed to be distributed normally with mean of zero and variance of σ_i^2. We estimate $\hat{g}_i$ and $\hat{r}_i$ using nonlinear tobit maximum likelihood, and recover estimated values of the underlying CES parameters $\hat{\alpha}_i$ and $\hat{\rho}_i$ from these.[4]

4.3.3 Predicting Conditional Cooperation

In sequential prisoner's dilemmas, there are two players - the first-mover and the second-mover - who sequentially choose whether to cooperate or defect. The extensive form of the sequential prisoner's dilemma where payoffs are symmetric is shown in Figure 4.1.

Figure 4.1: The Sequential Prisoner's Dilemma (SPD)



Note: $T > R > P > S$ and $2R > T + S$

The first-mover chooses to either cooperate (C) or defect (D), and, after observing this choice, the second-mover then chooses to cooperate (c) or defect (d). Subjects' earnings are maximized when both cooperate and each player

[4] We use the Matlab code (https://www.aeaweb.org/articles?id=10.1257/aer.97.5.1858) of Fisman et al. (2007) for the estimation of individual-level distributional preferences, slightly modified to impose a constraint so that $\hat{\rho}$ does not exceed one.

receives R. However, if the first-mover cooperates, the second-mover maximizes her own earnings by defecting, in which case the first-mover receives S and the second-mover receives T. If the first-mover defects, the second-mover maximizes her own earnings by also defecting, so that each player receives P. If players are own earnings maximizers, and this is common knowledge, there is a unique equilibrium where the first-mover defects and the second-mover defects regardless of the first-mover's choice.

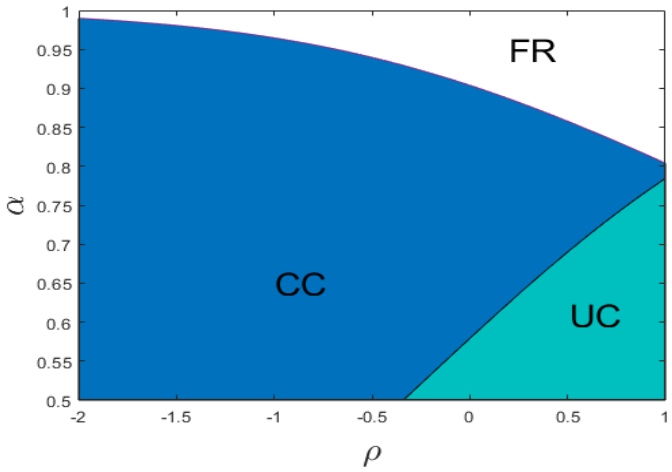
The CES model predicts that i) the second-mover cooperates in response to first-mover's cooperation if and only if the utility of mutual cooperation outweighs that of defecting against cooperator (i.e. $R^\rho > [\alpha T^\rho + (1-\alpha)S^\rho]$), and ii) the second-mover cooperates in response to first-mover's defection if and only if the utility of cooperating against defector outweighs that of mutual defection (i.e. $[\alpha S^\rho + (1-\alpha)T^\rho] > P^\rho$).

Therefore, the second-mover's utility maximizing strategy depends on preferences (α, ρ) and payoffs (T, R, P, S) as follows. If $\alpha > max\{\frac{R^\rho - S^\rho}{T^\rho - S^\rho}, \frac{T^\rho - P^\rho}{T^\rho - S^\rho}\}$, the optimal strategy is to defect regardless of the first-mover's choice (free-riding, henceforth FR). If $\alpha < min\{\frac{R^\rho - S^\rho}{T^\rho - S^\rho}, \frac{T^\rho - P^\rho}{T^\rho - S^\rho}\}$, the optimal strategy is to cooperate regardless of the first-mover's choice (unconditional cooperation, henceforth UC). If $\frac{R^\rho - S^\rho}{T^\rho - S^\rho} < \alpha < \frac{T^\rho - P^\rho}{T^\rho - S^\rho}$, the optimal strategy is to cooperate if the first-mover defects and defect if the first-mover cooperates (mis-matching , heceforth MM). Finally, if $\frac{T^\rho - P^\rho}{T^\rho - S^\rho} < \alpha < \frac{R^\rho - S^\rho}{T^\rho - S^\rho}$, the optimal strategy is to defect if the first-mover defects and cooperate if the first-mover cooperates (conditional cooperation, henceforth CC).[5]

[5] The likelihood of conditional cooperation is increasing in R and P and decreasing in T and S. In terms of the indices we will be using (and which will be explained in more detail in Section 4.4), the likelihood of conditional cooperation is increasing in DAMAGE ($\frac{R-S}{R}$) and decreasing in GAIN ($\frac{T-R}{R}$). For our parameterization, where R is constant across games, the likelihood of conditional cooperation is also decreasing in EFF ($\frac{R-P}{R}$).

Thus, for any given set of payoffs, estimates of α and ρ allow us to predict the second-mover's strategy. For example, suppose payoffs are given as in $T = 600, R = 500, P = 200, S = 90$. Figure 4.2 shows the second-mover's predicted strategy in this example. If $\alpha = \rho = 0.8$ then the player is predicted to conditionally cooperate, whereas if $\alpha = \rho = 0.9$ the player is predicted to free-ride.

Figure 4.2: An Example of the Second-mover's Predicted Strategy in the CES model



Note: CC: Conditional Cooperation, FR: Free-Riding, UC: Unconditional Cooperation.

To compare predictions with actual strategies we need to observe how a second-mover would respond to cooperation and how they would respond to defection. In this study, we elicit second-mover strategies using a strategy method (Selten, 1967) whereby subjects submit second-mover strategies in an incentive compatible way. We do this for twenty-four sequential prisoner's dilemma games with varying payoffs.[6]

[6] In Appendix A, we show how the preference space of (α, ρ) for second-mover strategies differs across twenty-four games.

4.4 Experimental Design and Procedures

The experiment was conducted at the University of Nottingham in November and December 2019. Subjects were recruited from a university-wide pool of students who had previously indicated their willingness to be paid volunteers in decision-making experiments.[7] The experiment consisted of two parts, one-week apart. Before subjects sign-up for the experiment, subjects were informed of the following information: 1) the experiment consists of two parts, one-week apart, 2) to participate in the experiment, subjects are required to attend both parts, 3) subjects will be paid in private and in cash at the end of part 2, and 4) each part will take approximately 90 minutes. We recruited 123 subjects for four sessions: $30 \sim 32$ subjects per session.

Upon arrival, subjects were welcomed and randomly seated at separated computer terminals. Subjects were then given a written set of instructions, which the experimenter read aloud. Then subjects were asked to solve a set of control questions to make sure they understand how their choices determine their earnings. Subjects had to answer all questions correctly before the experiment could continue. Once every participant correctly answered the control questions, subjects practiced an example task twice to familiarize them with the task and to fully understand the experimental procedure.

In part 1 (week 1), we used modified dictator games as developed by Fisman et al. (2007). Subjects were asked to allocate tokens between him/herself (π_s) and others (π_o) in fifty independent decision tasks. Each task is presented graphically on a computer screen: tokens for him/herself is given at a vertical

[7] Subjects were recruited through the online recruitment system ORSEE (Greiner, 2015). The experiment was programmed and conducted with the software LIONESS (Giamattei, Yahosseini, Gächter and Molleman, 2020). Full experimental instructions and experiment screenshots are provided in Appendix C.

axis and tokens for other is given at a horizontal axis. For each task, the computer selected a budget line at random from the set of lines that intersect at least one of the axes at 50 or more than 50 experimental currency tokens, but with no intercept exceeding 100 tokens or below 0 tokens. Thus, income (m) and relative price of giving (p_o/p_s) vary across decision situations allowing us to identify preference parameters.

Subjects were asked to allocate tokens by using the computer mouse or keyboard arrows to move the pointer to the desired allocation (π_s, π_o) on the budget line, and then clicking the submit button to confirm their choice. Once every subject starts the task, the computer randomly decides the relative price of giving and the amount of income in each game at the individual level. Subjects did not receive any feedback on the other's choice or the outcome of each modified dictator game until the end of part 2 (week 2) experiment.

Decisions in part 1 were incentivized as follows. At the end of part 2 (week 2), a number from 1 to 50 is randomly chosen to determine which decision situation out of fifty modified dictator games is used to determine the additional earnings form part 1. For this situation, each subject receives the sum of the tokens 1) he/she allocated to him/herself, and 2) a matched participant allocated to other. To prevent potential reciprocation, subjects were always randomly matched with two other subjects - one who will receive tokens from his/her own decision situation, and another who will give tokens to him/her. The tokens were converted to cash at the rate of 10 tokens $=£$ 1.

As we expected, there was some attrition in our experiment: 13% (16 out of 123) part 1 subjects did not participate in part 2. Because we needed an even number of participants to match subjects in sequential prisoner's dilemmas, we recruited reserves for the part 2 sessions. At the time of recruitment, they

were informed that the experiment will take 90 minutes and they will be paid in private in cash at the end of the experiment. We conducted four sessions in part 2 with 107 subjects who had participated in part 1 and 7 reserves.

In part 2 (week 2), we used sequential two-player prisoner's dilemma (henceforth SPD) games where the first-mover decides first and the second-mover decides after observing the first-mover's decision. In each game subjects had to make decisions in both the role of first-mover and the role of second-mover.

Each subject played twenty-four SPDs: eight symmetric and sixteen asymmetric games. For the symmetric games, we used the payoff parameterization from Gächter, Lee and Sefton (2020), who orthogonally vary the indices of risk (RISK, $\frac{P-S}{P}$), temptation (TEMPT, $\frac{T-R}{T}$), and efficiency (EFF, $\frac{R-P}{R}$) to study cooperation in simultaneous prisoner's dilemmas. For each of these games we created two asymmetric games : one (which we refer to as the Ahead game) by adding $T-S$ to the player's own payoffs, and one by adding $T-S$ to the opponent's payoffs (which we refer to as the Behind game).[8] Therefore, our experimental design allows us to examine how distributional preferences shape second-mover strategy in both symmetric and asymmetric games across a broad set of payoff parameterizations.[9]

Table 4.1 shows the payoff parameterizations for twenty-four sequential prisoner's dilemma games: symmetric (eight games: S1-S8) and asymmetric SPDs (eight games with advantageous payoffs: A1-A8, eight games with disadvantageous payoffs: B1-B8).

[8] Note that in the Ahead game the player's payoff is always as much as the opponent's, even if the player cooperates and the opponent defects.

[9] Note that asymmetric games allow us to discriminate between the predictions of the CES model and the Fehr-Schmidt model. For example, in the Behind games the CES model predicts that conditional cooperation could arise, whereas the Fehr-Schmidt model predicts unilateral defection regardless of the first-mover's choice.

Table 4.1: Payoff Parameterization for Sequential Prisoner's Dilemmas

Game	T	R	P	S	DAMAGE	GAIN	EFF
S1	600	500	200	90	0.82	0.20	0.60
S2	600	500	200	20	0.96	0.20	0.60
S3	800	500	200	90	0.82	0.60	0.60
S4	800	500	200	20	0.96	0.60	0.60
S5	600	500	400	180	0.64	0.20	0.20
S6	600	500	400	40	0.92	0.20	0.20
S7	800	500	400	180	0.64	0.60	0.20
S8	800	500	400	40	0.92	0.60	0.20
A1	1110	1010	710	600	0.82	0.10	0.30
A2	1180	1080	780	600	0.96	0.09	0.28
A3	1510	1210	910	800	0.82	0.25	0.25
A4	1580	1280	980	800	0.96	0.23	0.23
A5	1020	920	820	600	0.64	0.11	0.11
A6	1160	1060	960	600	0.92	0.09	0.09
A7	1420	1120	1020	800	0.64	0.27	0.09
A8	1560	1260	1160	800	0.92	0.24	0.08
B1	600	500	200	90	0.41	0.20	0.60
B2	600	500	200	20	0.44	0.20	0.60
B3	800	500	200	90	0.34	0.60	0.60
B4	800	500	200	20	0.38	0.60	0.60
B5	600	500	400	180	0.35	0.20	0.20
B6	600	500	400	40	0.43	0.20	0.20
B7	800	500	400	180	0.29	0.60	0.20
B8	800	500	400	40	0.37	0.60	0.20

Note: S denotes 'Symmetric' games, A denotes 'Ahead' games (i.e. asymmetric games with advantageous payoffs), and B denotes 'Behind' (i.e. asymmetric games with disadvantageous payoffs). In the Ahead (Behind) games, the opponent's payoffs are equivalent to the payoffs of the Behind (Ahead) games. Note that, taking the T, R, P and S parameters specified for the asymmetric games as those of the second-mover and taking that player's perspective, DAMAGE depends on the opponent's payoff parameters, while GAIN and EFF depend on own payoff parameters. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$.

The indices of RISK, TEMPT, and EFF capture important motivations that are relevant in a simultaneous prisoner's dilemma or in the first-mover's decision, where players face uncertainty about the choice of the opponent. In the SPDs, second-movers do not face such uncertainty and the second-mover's decision directly determines his/her own earnings and the opponent's earnings.

Therefore, we employ the indices of DAMAGE and GAIN that are more closely linked to second-mover's decision. DAMAGE ($\frac{R-S}{R}$) refers to the losses imposed on the first-mover when the second-mover defects against cooperation, while GAIN ($\frac{T-R}{R}$) represents the gains to the second-mover from defecting against to the first-mover's cooperation. We also employ the index of EFF ($\frac{R-P}{R}$), which refers to the gains of mutual cooperation compared to mutual defection. Holding R as constant, DAMAGE, GAIN, and EFF captures the effect of S, T, and P on second-mover strategies, respectively.

The implied values of DAMAGE, GAIN, and EFF are as follows. The value 0.64 of DAMAGE implies that an individual who defects against a cooperator will reduce the first-mover's monetary payoff by 64% compared to the payoff from mutual cooperation. The value 0.20 of GAIN implies that an individual who defects against a cooperator will accrue a 20% gain in monetary payoff compared to the payoff from mutual cooperation. The value 0.60 of EFF implies that mutual cooperation will yield a 60% gain in monetary payoff compared to the payoff from mutual defection.

Subjects were randomly paired and then played all twenty-four games in random order at the pair level with no feedback between games. Both decision tasks were presented on the same screen. In the *First-mover's decision*, they simply chose whether to cooperate or to defect as a first-mover. In the *Second-mover's decision*, we asked subjects to decide in the following two situations:

i) if the first-mover cooperates, and ii) if the first-mover defects. Therefore, we elicited second-mover's decision using the strategy method (Selten, 1967).[10] Decisions were neutrally labelled as options A and B, in which "A" is cooperation and "B" is defection.

Decisions in part 2 were incentivized as follows. After all pairs had completed all twenty-four games, the computer randomly selected one out of twenty-four games at the pair level. Then, one of the pair was randomly chosen to be the first-mover and the other to be the second-mover. The first-mover's decision was matched with the decision of the second-mover, and these determined the additional earnings from part 2. Subjects were then informed of which game was chosen for their payoff-relevant situation and the outcome in that game. The tokens were converted to cash at the rate of 100 tokens = £ 1.

At the end of part 2, subjects were asked to fill out a post-experimental questionnaire asking for demographic and attitudinal information. Demographic information includes age, gender, ethnicity, major, expenditure per month (excluding rent), and a self-assessment of subjects' political orientation.[11] Attitudinal information included a self-assessment of risk and time preferences, positive and negative reciprocity, trust, and self-assessment of math skills and procrastination using the questionnaire developed by Falk et al. (2018), and self-evaluation of personality using Big Five personality test suggested by Hahn et al. (2012).

[10] Regarding potential differences between responses elicited using the strategy method and those using a direct response method, previous studies found no statistical differences in subjects' responses between these two methods (see Brandts and Charness, 2000; 2011 for a review).

[11] To elicit political orientation, we use an 11-point left-right scale that is frequently used in surveys (e.g., Costa Lobo and Curtice, 2015; Kroh, 2007).

Overall, the total earnings of a subject are the sum of the participation fees (£ 3 per each part), and the additional earnings from parts 1 and 2. Both parts lasted approximately 90 minutes and, for subjects who completed both parts of the experiment, subject's earnings ranged from £ 8 to £ 34.30, averaging £ 17.80.

4.5 Results

The following analysis is structured to discuss our main research questions. First, we overview individual-level distributional preferences of ($\hat{\alpha}$, $\hat{\rho}$), which are estimated from the allocation decisions in the modified dictator games. Second, we generate predictions of second-mover's choices in twenty-four SPDs based on individual estimates of distributional preferences. We compare the predictions with actual choices to investigate the predictive power of estimated distributional preferences. Lastly, we use regression analysis to examine the relationship between estimated preference parameters and behavior in SPDs.

For our analysis, we only include the decisions of subjects who completed both parts of the experiment: thus our data consists of 107 subjects' decisions. For Part 1 (modified dictator game) and Part 2 (sequential prisoner's dilemma game), we have 5,350 (107 subjects × 50 games) and 2,568 (107 subjects × 24 games) observations, respectively.

4.5.1 Estimated Distributional Preferences

Before we estimate individual-level distributional preferences, we examine whether subjects' allocation decisions in the modified dictator games are consistent with utility maximization. We use Afriat's (1972) Critical Cost Efficiency Index (CCEI) to measure consistency with the Generalized Axiom of

Revealed Preference (GARP). CCEI measures the amount by which each budget constraint should be adjusted in order to remove all violations of GARP.[12]

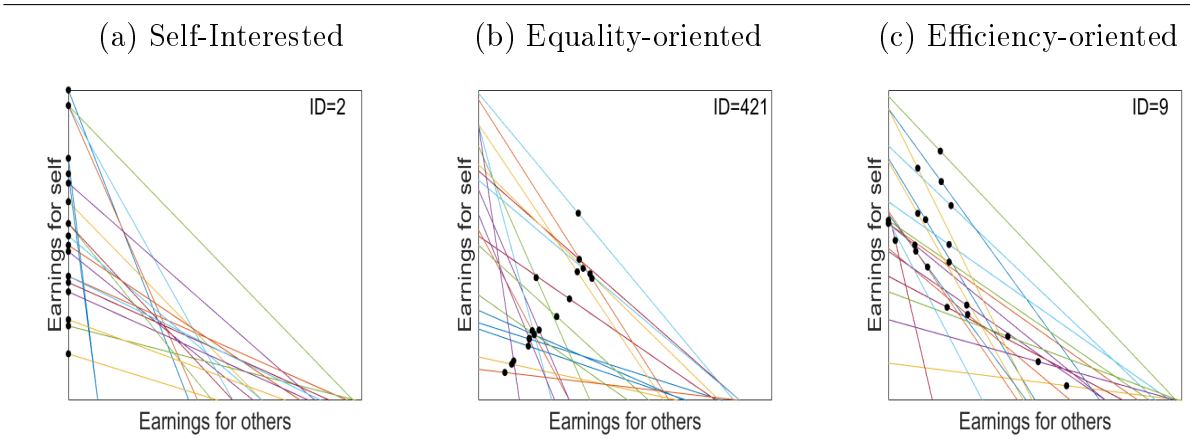
We find that 80% of our subjects have a CCEI > 0.80. By way of comparison, 86% of subjects in the lab experiment of Fisman, Kariv and Markovits (2007) exhibit a CCEI score higher than 0.80, and 74% of subjects in the online experiment of Fisman, Jakiela and Kariv (2017) exhibit a CCEI score higher than 0.80.[13]

There are substantial heterogeneities across subjects' allocations (π_s, π_o), where π_s represents their own payoffs and π_o denotes other's payoffs they are randomly matched with. Figure 4.3 illustrates three examples of actual subject choices in the modified dictator games. Figure 4.3(a) describes a self-interested subject's choices: the subject allocates all tokens for him/herself in most games regardless of variations in the relative price of giving. Figure 4.3(b) presents a subject's choices that reveal a taste for equality: the subject allocates a similar amount of tokens to self and other even when the relative price of giving becomes expensive. Figure 4.3(c) shows a subject's choices that exhibit a taste for efficiency: the subject allocates more tokens to others when the relative price of giving becomes cheaper, but allocates less tokens to others when the relative price of giving becomes expensive.

[12] The CCEI ranges from zero to one. The CCEI closer to one implies the smaller adjustment of the budget constraints is required to eliminate all GARP violations.

[13] CCEI scores tend to decrease as the number of budget sets increases. Andreoni and Miller (2002) finds that almost all subjects' CCEI scores are close to one from eight-round modified dictator games, whereas Fisman et al. (2007, 2017) reports relatively lower CCEI scores from fifty-round modified dictator games. Choi et al. (2007) suggests the importance of collecting choice data from a wide range of budget sets to rigorously test utility maximization by showing that a random choice generator will pass the GARP test with high probability when the number of individual-level choices is small.

Figure 4.3: Three Examples of Subjects' Choices in the Modified Dictator Games

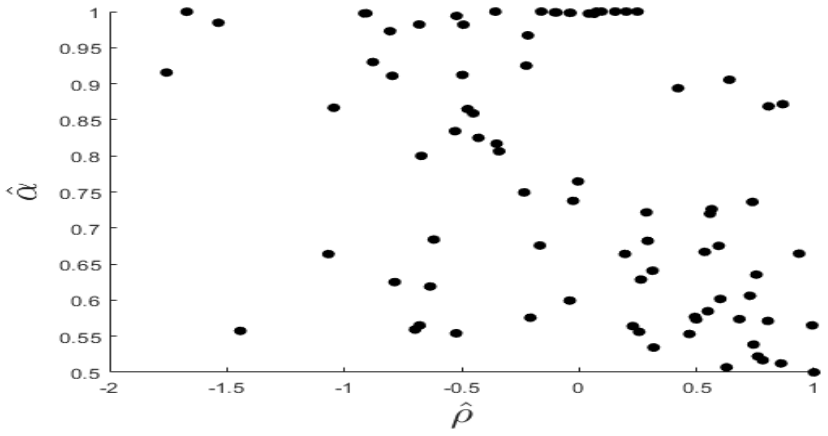


Among 107 subjects, 29 subjects (27%) never allocate more tokens to others than themselves. The rest of 78 subjects allocate more tokens to others than themselves at least once in fifty modified dictator games: 37 subjects (34.6%), 23 subjects (21.5%) and 14 subjects (13.1%) allocate more tokens to others in up to 10 games, 11 $\sim$ 20 games, and 21 $\sim$ 30 games, respectively. These subjects allocate more tokens to others when the relative price of giving is relatively inexpensive. For those games in which subjects allocate more tokens to others than themselves, the relative price of giving ($\frac{p_o}{p_s}$) is less than one in most of those games (81.6%): $\frac{p_o}{p_s} < 1$ implies that subjects can give more than one unit of earnings to others by decreasing their own earnings by one unit.

Figure 4.4 presents individual distributional preference parameter estimates for each subject. Among 107 subjects, 31 subjects (29%) can be classified as (weakly) self-interested ($\hat{\alpha} > 0.95$); 39 subjects (36%) reveal a taste for reducing inequality in payoffs ($\hat{\alpha} \leq 0.95$ and $\hat{\rho} < 0$), while 37 subjects (35%) exhibit a taste for increasing total payoffs ($\hat{\alpha} \leq 0.95$ and $\hat{\rho} > 0$).[14]

[14] For the classification of (weakly) self-interested, equality-oriented, and efficiency-oriented, we refer to Fisman, Jakiela and Kariv (2017).

Figure 4.4: Individual-level Estimates of $\hat{\alpha}$ and $\hat{\rho}$



4.5.2 Second-mover's Strategy

All subjects submit a strategy as a second-mover for each SPD. The strategy can specify either conditional cooperation (cooperate if and only if the first-mover cooperates, CC), free-riding (defect regardless of first-mover's choice, FR), unconditional cooperation (cooperate regardless of first-mover's choice, UC), or mismatching (cooperate if the first-mover defects and defect if the first-mover cooperates, MM).

Table 4.2 describes the proportions of second-mover strategies in symmetric and asymmetric games. Overall, CC and FR strategies predominate among the four strategies: 21% and 67% of strategies are CC and FR.

Table 4.2: Proportion of Second-mover Strategies

	Symmetric	Ahead	Behind	Average
CC	0.336	0.154	0.146	0.212
FR	0.605	0.617	0.778	0.667
UC	0.035	0.167	0.039	0.080
MM	0.023	0.062	0.037	0.041

Note: CC: Conditional Cooperation, FR: Free-Riding, UC: Unconditional Cooperation, MM: Mis-Matching. Ahead denotes asymmetric sequential prisoner's dilemmas with advantageous payoffs, Symmetric represents sequential prisoner's dilemmas with symmetric payoffs, and Behind indicates asymmetric sequential prisoner's dilemmas with disadvantageous payoffs.

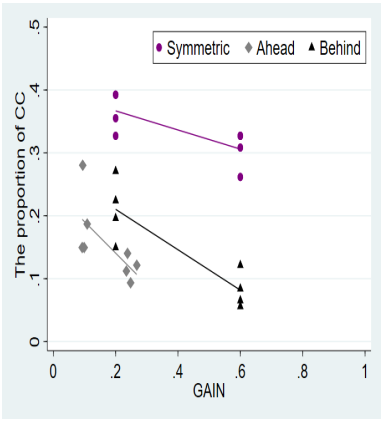
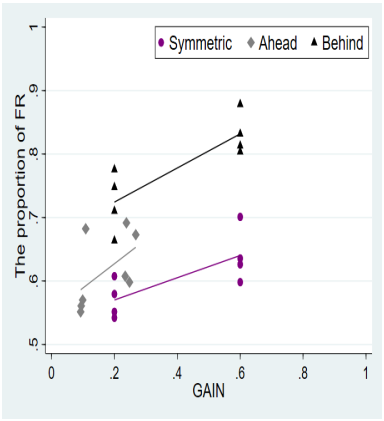
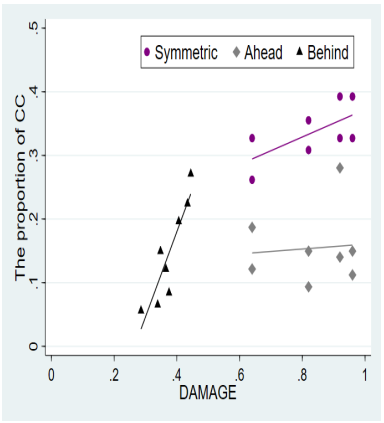
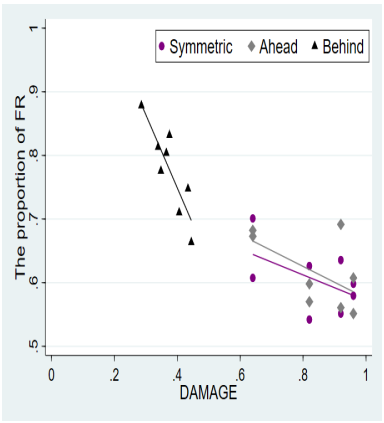
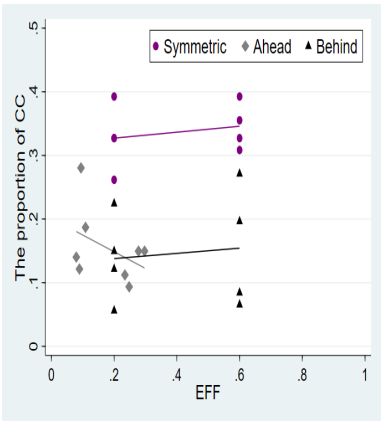
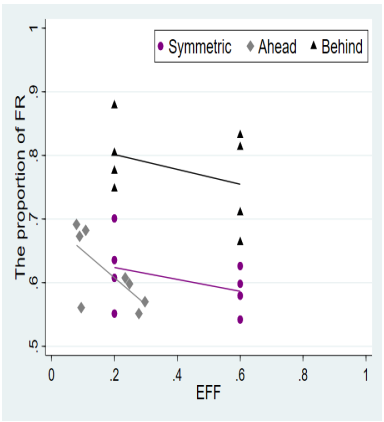
We find two noteworthy patterns from the proportion of second-mover strategy across games. First, free-riding notably increases in asymmetric games with disadvantageous payoffs. The proportion of FR is 78% in games with disadvantageous payoffs, compared with 60% in symmetric games and 62% in games with advantageous payoffs. Second, unconditional cooperation occurs more frequently in asymmetric games with advantageous payoffs than other games. When subjects have advantageous payoffs, the average proportion of UC is 17%, wheras it is less than 4% in other games.

Figure 4.5 illustrates that the proportions of CC and FR strategies vary with GAIN, DAMAGE, and EFF. In both symmetric and asymmetric games, the proportions of CC strategies decrease with GAIN and increase with DAMAGE, whereas the proportions of FR strategies increase with GAIN and decrease with DAMAGE.[15] We do not observe a clear relation between the proportions of CC strategies and EFF, whereas the proportions of FR strategies decrease with EFF in both symmetric and asymmetric games.[16]

[15] The proportion of CC strategies exhibits a strong negative correlation with GAIN (i.e. the Spearman correlation, symmetric : -0.788, ahead: -0.743, behind: -0.873); and the proportion of FR strategies is strongly positively correlated with GAIN (i.e. symmetric: 0.764, ahead: 0.691, behind: 0.873) as well. DAMAGE exhibits a strong correlation with the proportion of CC strategies in behind games (the Spearman correlation, 0.905), but moderate correlation in symmetric games (0.579), and a weak correlation in ahead games (-0.049). The proportion of FR strategies also shows similar patterns with DAMAGE (i.e. symmetric: -0.390, ahead: -0.439, behind: 0.809).

[16] The proportion of CC strategies exhibit a weak correlation with EFF (i.e. the Spearman correlation, symmetric : 0.169, ahead: -0.036, behind: 0.109). EFF exhibits moderate negative correlations with the proportion of FR strategies in symmetric (-0.436) and ahead games (-0.667) but a weak negative correlation in behind games (-0.218).

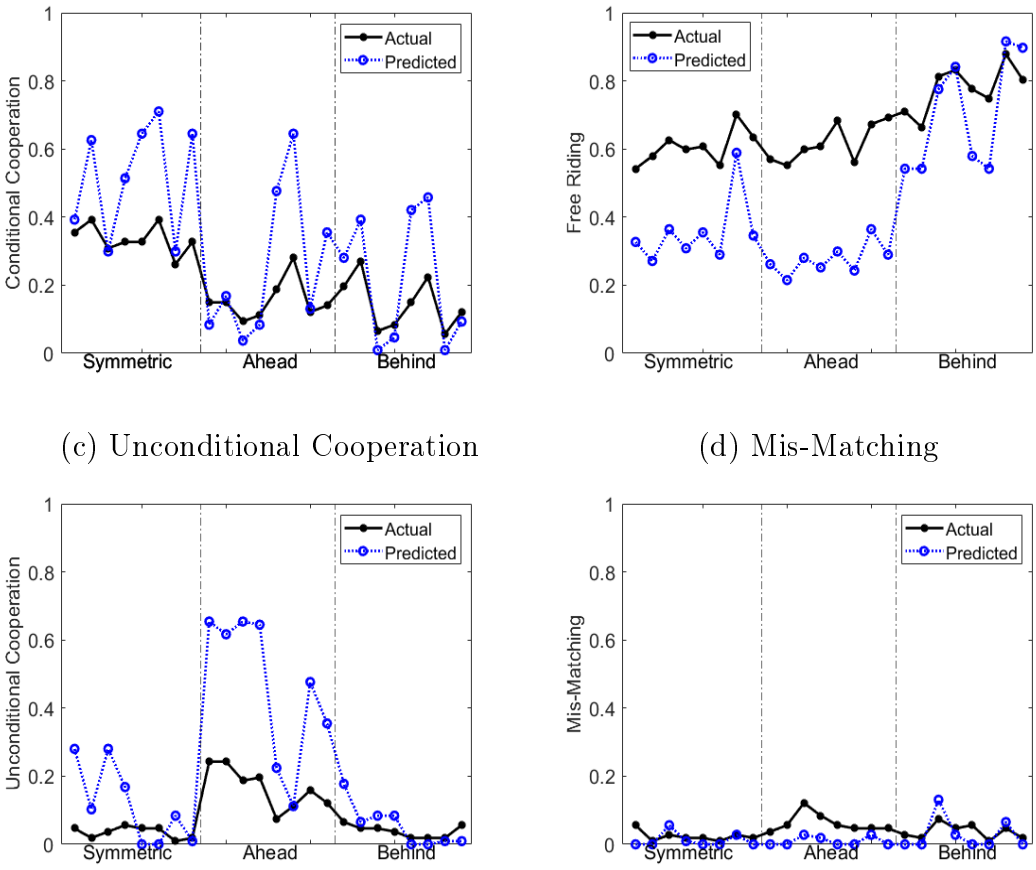
Figure 4.5: Proportions of Conditional Cooperation and Free Riding

	(a) Conditional Cooperation (CC)	(b) Free Riding (FR)
GAIN		
DAMAGE		
EFF		

4.5.3 Comparing Actual and Predicted Strategies

Based on individual-level estimates of $(\hat{\alpha}, \hat{\rho})$, we generate the predicted second-mover strategy of each subject in twenty-four sequential PDs. We first assess the accuracy of predictions at the aggregate level. Figure 4.6 compares the predicted and actual second-mover's strategies at the aggregate level.

Figure 4.6: Proportions of Second-mover Strategies: Actual and Predicted

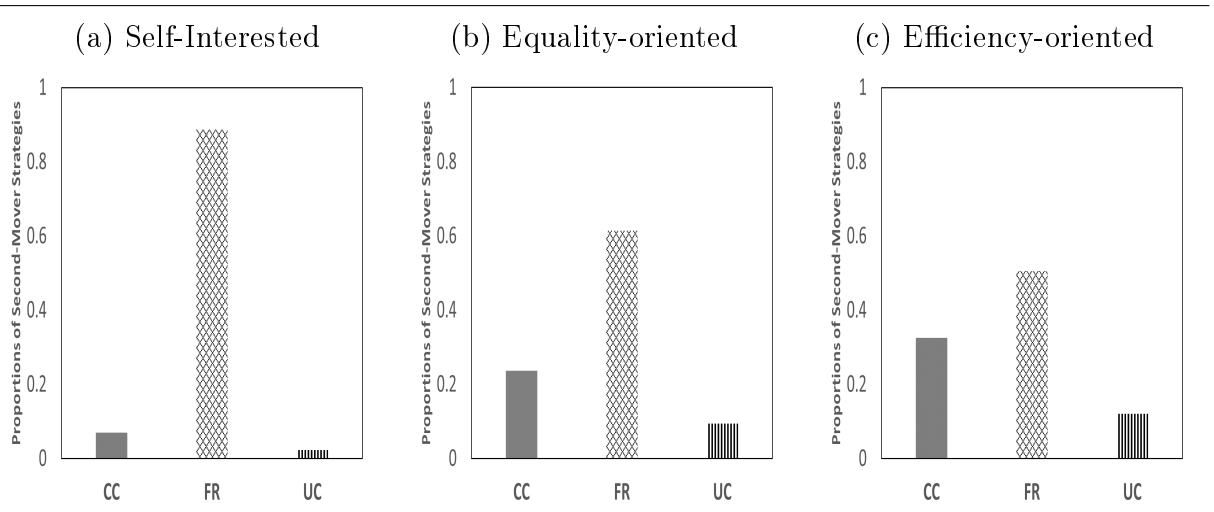


There exist substantial differences in levels between the predicted and actual choices: In most games, the proportions of conditional cooperation and unconditional cooperation are lower than the predictions, whereas free-riding occurs more frequently than the prediction. Note, however, that there is a high

Spearman correlation between the predicted and actual proportion of second-mover strategy for conditional cooperation ($r_s = 0.868$), free-riding (0.841), and unconditional cooperation (0.748), but not mismatching (0.364).

Figure 4.7 presents the proportions of second-mover strategies across subjects' types: self-interested ($\hat{\alpha} > 0.95$), equality-oriented ($\hat{\alpha} \leq 0.95$ and $\hat{\rho} < 0$), and efficiency-oriented ($\hat{\alpha} \leq 0.95$ and $\hat{\rho} > 0$). Self-interested subjects (Figure 4.7(a)) mostly free-ride and the proportions of CC and UC are marginal. In contrast, for both equality-oriented (Figure 4.7(b)) and efficiency-oriented (Figure 4.7(c)) subjects, the proportions of CC and UC notably increase while the proportion of FR substantially decreases. This pattern is most clear in the symmetric games: the proportions of CC for equality-oriented and efficiency-oriented subjects are 39% and 49%, respectively, whereas the proportion of CC for self-interested subjects is only 8%.

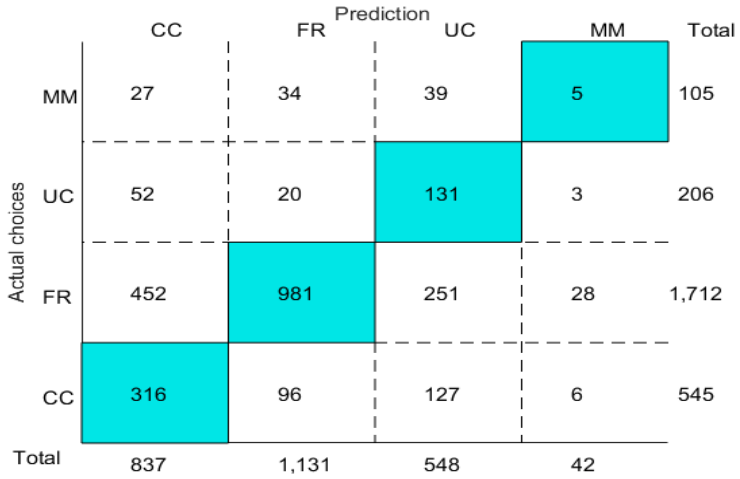
Figure 4.7: Proportions of Second-mover Strategies across Subjects' Types



Note: CC: Conditional Cooperation, FR: Free-Riding, UC: Unconditional Cooperation.

We examine the accuracy of predictions at the *aggregate* level. As shown in Figure 4.8, our data set consists of 2,568 strategies (107 subjects × 24 SPDs). The CES model correctly predicts 1,433 out of 2,568 strategies (i.e., 55.8%).

Figure 4.8: Comparison between Predicted and Actual Choices



First, we find that the CES model overpredicts conditional cooperation and underpredicts free-riding. The model predicts that conditional cooperation occurs 837 cases but the actual number of conditional cooperation is 545 cases. In contrast, the model predicts that free-riding occurs 1,131 cases but the actual number of free-riding is 1,712 cases.

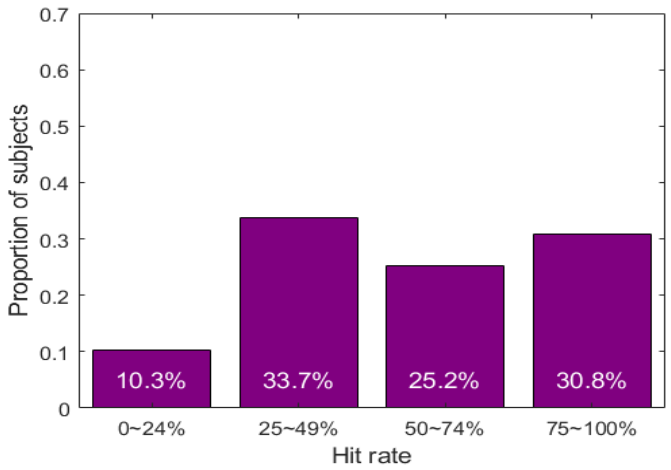
Second, the CES model shows better performance in predicting FR than in predicting CC. For FR, 86.7% (981 out of 1,131) of predictions coincide with the actual choices. For CC, however, only 37.8% (316 out of 837) of predictions corresponds to the actual choices.

These two findings are similar to Blanco et al. (2011). In the experiment of Blanco et al., the Fehr-Schmidt model overpredicts conditional cooperation (i.e. 41 cases are predicted but the actual number of CC is 20) while it underpredicts free-riding (i.e. 20 cases are predicted but the actual number of

FR is 37). Blanco et al. find that the Fehr-Schmidt model shows higher predictive accuracy when the model predicts free-riding. For free-riding, 15 out of 20 (75.0%) predictions correspond to the actual choices. In contrast, for conditional cooperation, only 17 out of 41 (41.4%) predictions coincide with the actual choices.

Next, we examine the accuracy of predictions at the *individual* level. Figure 4.9 reports the proportions of subjects with varying hit rates. Out of 107 subjects, 60 (56.1%) have a hit rate of 50% or higher. That is, for these subjects their strategy choices correspond to predictions in at least 12 of 24 games. Among these 60 subjects, 27 subjects are self-interested, 15 subjects are efficiency-oriented, and 18 subjects are equality-oriented. Using a stricter threshold of prediction accuracy of 75% (i.e., 18 out of 24 games), approximately one-third of subjects' (33 out of 107 subjects) strategies coincide with predictions. Among these 33 subjects, 23 subjects are self-interested, while 6 and 4 subjects exhibit preferences toward efficiency and equality, respectively.

Figure 4.9: The Hit Rate of the CES model's Prediction at the Individual Level



Note: Hit rate = (a number of games where prediction matches with actual choice)/24

4.5.4 Comparing with Alternative Benchmarks

We compare the predictive power of the CES model using the following benchmarks: a random choice generator, the Fehr-Schmidt model, the Charness-Rabin model, and the self-interested model. A random-choice generator denotes where agents randomly choose either CC, FR, UC, or MM with a probability of 25% in a sequence of 24 independent games. For the Fehr-Schmidt model, we estimate the parameters of advantageous inequality aversion (β) from the modified dictator games. For the Charness-Rabin model, we estimate the weight on own income (λ) and the degree of concern for equality (δ) versus maximizing the total income from the modified dictator games. The self-interested model assumes that agents only care for their own earnings. The details of the Fehr-Schmidt and the Charness-Rabin models and their estimation are provided in Appendix A.

Table 4.3 compares the proportion of subjects with a hit rate equal or greater of 50% for each model. The first row compares the performance of the models for all twenty-four games, the second row compares the performance of the models for the symmetric games. The third and fourth rows compare the performance of the models for the Ahead and Behind games, respectively.

Table 4.3: The Proportions of Subjects with a Hit Rate equal or greater of 50%

	CES	RC	FS	CR	SF
All	0.561	0.007	0.598	0.542	0.645
Symmetric	0.533	0.114	0.626	0.570	0.608
Ahead	0.477	0.114	0.439	0.411	0.626
Behind	0.766	0.114	0.832	0.794	0.832

Note: CES - Constant Elasticity of Substitution, RC - a Random Choice benchmark, FS - the Fehr-Schmidt model, CR - the Charness-Rabin model, SF - the self-interested model.

First, all models improve considerably over the random choice benchmark. That benchmark would be expected to predict the correct strategy in at least

half of a subjects' game for only 0.7% of subjects, while across all games, each model predicts most strategies correctly for the majority of subjects.

Second, the distributional preference models, - the CES, the Fehr-Schmidt, and the Charness-Rabin models -, show similar performance in predicting second-mover strategies. Across all twenty-four games, we observe that the Fehr-Schmidt model slightly performs better than the CES and the Charness-Rabin models, but the differences are quite small: based on the Fehr-Schmidt model predictions, 59.8% of subjects have a hit rate of 50% or higher, while the CES model and the Charness-Rabin models correctly predict 56.1% and 54.2% of subjects' second-mover strategies at a hit rate of 50% or higher.

Third, the self-interested model shows better performance than other models. Based on the self-interested model predictions, 64.5% of subjects have a hit rate of 50% or higher in twenty-four games. This is because non-selfish subjects free-ride more often than distributional preferences predict: their strategies correspond to the predictions based on their distributional preferences in some games, while they free-ride in other games in which they are predicted to choose cooperative strategies (i.e. (un)conditional cooperation).

Lastly, the performances of the models differ across games. In the symmetric games, the Fehr-Schmidt model shows better performance than other models: in terms of the proportion of subjects with a hit rate of 50% or greater, the Fehr-Schmidt model comes out first, followed by the self-interested, the Charness-Rabin, and the CES models. In the Ahead games, the self-interested model comes out first, followed by the CES, the Fehr-Schmidt, and the Charness-Rabin models. In the Behind games, the Fehr-Schmidt and the self-interested models come out first, followed by the Charness-Rabin, and the CES models.

4.5.5 Regression Analysis

In this subsection, we conduct a regression analysis to address the following question: How are distributional preferences linked to second-mover strategy?

Table 4.4 shows the results of random effect probit regressions that examine the determinants of conditional cooperation in the symmetric games. Column (1) examines how conditional cooperation varies with the payoff indices (i.e. DAMAGE, GAIN, and EFF). Column (2) adds controls for individual characteristics and session effects in addition to the regressors of column (1). Column (3) adds individual-level parameters of distributional preferences $(\hat{\alpha}, \hat{\rho})$ in addition to the regressors of column (2).

First, the indices of DAMAGE and GAIN have a significant effect on conditional cooperation in all columns (1)-(3). Subjects are more likely to conditionally cooperate when free-riding has a larger negative impact on the first-mover's earnings or when subjects would gain less additional earnings from free-riding. Based on the complete sample, the probability of conditional cooperation increases by $2.1 \sim 2.2$ percentage points when DAMAGE increases by 0.1 point, while the probability of conditional cooperation decreases by 1.4 percentage points when GAIN increases by 0.1 point. The index of efficiency, however, does not have a significant effect on conditional cooperation.[17]

Second, some individual characteristics are influential determinants of conditional cooperation. Among attitudinal characteristics, only Falk et al.'s (2018) measure of positive reciprocity has a significant effect on conditional cooperation: subjects who have a strong desire to reciprocate to other's kind

[17] Chapter 3, which implemented an online sequential prisoner's dilemma experiment to examine how within-subject conditional cooperation is influenced by payoff variations, finds similar results. We find that DAMAGE and GAIN have a significant effect on conditional cooperation, whereas we observe that EFF has an insignificant effect.

actions are more likely to conditionally cooperate. Among demographic traits, we find that major and political orientation have a significant effect: subjects who study economics are less likely to conditionally cooperate, while subjects who hold left-wing views are more likely to conditionally cooperate. We do not observe a significant effect of age and gender for conditional cooperation. Session effects are jointly insignificant in the complete sample (Wald test, $\chi^2 = 5.97$, $p = 0.1130$).

Table 4.4: Determinants of Conditional Cooperation: Symmetric games

	(1)	(2)	(3)
DAMAGE	0.216***	0.208***	0.206***
	(0.075)	(0.076)	(0.075)
GAIN	-0.140**	-0.137**	-0.136**
	(0.058)	(0.059)	(0.058)
EFF	-0.008	-0.021	-0.022
	(0.066)	(0.066)	(0.066)
$\hat{\alpha}$			-0.849***
			(0.132)
$\hat{\rho}$			-0.002
			(0.002)
CCEI			-0.068
			(0.178)
Round	0.001	0.001	0.001
	(0.002)	(0.001)	(0.001)
P.Reciprocity		0.136***	0.105***
		(0.045)	(0.039)
Controls	*No*	*Yes*	*Yes*
Observations	856	832	832
Pseudo R^2	0.021	0.077	0.120
Log-likelihood	-349.4	-321.4	-306.6
BIC	739.4	837.7	828.4

Note: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, elicited using the questionnaire of Falk et al. (2018). Controls: the variables of demographic information, attitudinal characteristics, and session effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Third, we find that conditional cooperation is significantly influenced by the degree of fair-mindedness. The probability of conditional cooperation decreases with $\hat{\alpha}$, which implies that subjects are less likely to conditionally cooperate as they put more weight on their own material earnings. Based on the complete sample with controls (column (3)), the probability of conditional cooperation decreases by 8.5 percentage points as $\hat{\alpha}$ increases by 0.1 point.

Fourth, we do not observe a significant effect of equality-efficiency concerns ($\hat{\rho}$) on conditional cooperation. $\hat{\rho}$ has a marginal and insignificant impact on conditional cooperation. This suggests that whether subjects are oriented toward equality or efficiency is not the main determinant for conditional cooperation. Both equality-oriented ($\hat{\alpha} \leq 0.95$ and $\hat{\rho} < 0$) and efficiency-oriented ($\hat{\alpha} \leq 0.95$ and $\hat{\rho} > 0$) subjects are more likely to conditionally cooperate than self-interested ($\hat{\alpha} > 0.95$) subjects (see Figure 4.7), but the propensity to conditionally cooperate does not differ across equality-oriented and efficiency-oriented subjects.

Next, we focus on conditional cooperation in the asymmetric games. In Table 4.5, columns (1)-(3) present the determinants of conditional cooperation in the *Ahead* games (i.e. asymmetric games with advantageous payoffs), while columns (4)-(6) present the determinants of conditional cooperation in the *Behind* games (i.e. asymmetric games with disadvantageous payoffs).

First of all, we find that weaker fair-mindedness (1-$\hat{\alpha}$) significantly impedes conditional cooperation. As in the symmetric games, subjects who put more weight on their own material earnings are less likely to conditionally cooperate. Note, however, that the effect size of $\hat{\alpha}$ is smaller in the asymmetric games (65% smaller in the Ahead games and 42% smaller in the Behind games).

Second, we observe that equality-efficiency tradeoffs ($\hat{\rho}$) have a marginal

effect in the Behind games (average marginal effect of $\hat{\rho}$ = -0.001, implying that the probability of CC increases by 0.01 percentage points when $\hat{\rho}$ decreases by 0.1 point). The parameter of equality-efficiency tradeoffs ($\hat{\rho}$) is not significant in the Ahead games.

Table 4.5: Determinants of Conditional Cooperation: Asymmetric games

	Ahead			Behind		
	(1)	(2)	(3)	(4)	(5)	(6)
DAMAGE	0.099	0.074	0.075	0.895***	0.829***	0.832***
	(0.089)	(0.085)	(0.085)	(0.313)	(0.297)	(0.288)
GAIN	-0.563***	-0.594***	-0.585***	-0.151*	-0.158**	-0.150*
	(0.167)	(0.168)	(0.165)	(0.082)	(0.080)	(0.078)
EFF	-0.390***	-0.397***	-0.390***	-0.025	-0.019	-0.018
	(0.149)	(0.150)	(0.147)	(0.053)	(0.053)	(0.051)
$\hat{\alpha}$			-0.300***			-0.497***
			(0.082)			(0.075)
$\hat{\rho}$			0.001			-0.001**
			(0.001)			(0.001)
CCEI			0.106			0.017
			(0.121)			(0.094)
Round	-0.002	-0.003	-0.003	-0.003	-0.003*	-0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
P.Reciprocity		0.059***	0.043**		0.038	0.019
		(0.022)	(0.020)		(0.028)	(0.021)
Controls	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Observations	856	832	832	856	832	832
Pseudo R^2	0.032	0.090	0.111	0.079	0.143	0.204
Log-likelihood	-334.2	-305.6	-298.6	-290.9	-260.9	-242.5
BIC	709.0	806.2	812.5	622.3	716.8	700.2

Note: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, elicited using the questionnaire of Falk et al. (2018). Controls: the variables of demographic information, attitudinal characteristics, and session effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Third, the significance of payoff indices differs across symmetric, Ahead, and Behind games. In the symmetric games, both DAMAGE and GAIN are

highly significant while EFF is not. In the Ahead games, GAIN and EFF are highly significant while DAMAGE is not: i) GAIN harms conditional cooperation because subjects are more likely to free-ride as GAIN increases, and ii) conditional cooperation decreases with EFF because subjects are more likely to *unconditionally* cooperate in the Ahead games when EFF increases. In the Behind games, DAMAGE is highly significant (i.e. at the 1% level) while GAIN is marginally significant (i.e. 10% level), and EFF is insignificant.

Fourth, the effect sizes of individual characteristics are influenced by payoff asymmetry. Positive reciprocity, whether studying economics, and political orientation still affect conditional cooperation in the Ahead games, but the effect size is notably reduced compared to the symmetric games. In the Behind games, almost all individual characteristics are not significant at the 5% level.

Our main findings from the regression results are three fold. First, conditional cooperation is influenced by payoff variations. In particular, subjects are more likely to conditionally cooperate when DAMAGE is high or GAIN is low. Second, individual characteristics, particularly positive reciprocity, whether studying economics, and political orientation, also affect conditional cooperation. Third, the estimates of distributional preferences help explain conditional cooperation. Especially, the estimate of the fair-mindedness parameter of the CES model significantly affects conditional cooperation. Note that individual characteristics which are significant in the symmetric games become insignificant in the Behind games, whereas fair-mindedness is significant at the 1% level in all regressions.

Overall, our results suggest that distributional preferences are influential determinants for conditional cooperation. A point worthy of further discussion is the insignificant effect of equality-efficiency tradeoffs ($\hat{\rho}$) for condi-

tional cooperation. In Appendix B, we report the regression results using the individual-level estimates of the following distributional preference models: the Fehr-Schmidt (1999) and the Charness-Rabin models (2002). For the Fehr-Schmidt model, we use the advantageous inequality aversion parameter ($\hat{\beta}$). For the Charness-Rabin model, we use the weight on own income ($\hat{\lambda}$) and the degree of concern for equality ($\hat{\delta}$) versus maximizing the total income.

We find that subjects who are averse to advantageous inequality (i.e. higher $\hat{\beta}$) are more likely to conditionally cooperate. We also find that subjects who are equality-oriented (i.e. higher $\hat{\delta}$) are more likely to conditionally cooperate in the symmetric games, and subjects who put more weight on other's earnings ($\hat{\lambda}$) are more likely to conditionally cooperate in the asymmetric games. Therefore, whether equality-efficiency concerns have a significant relationship with conditional cooperation depends on which distributional preference model is chosen.

4.5.6 First-mover's Cooperation

In this subsection, we examine the role of distributional preferences for cooperation under strategic uncertainty. First-mover's cooperation depends on not only his/her own distributional preferences but also subjective beliefs about whether or not second-movers will cooperate. Although the relationship between first-mover's cooperation and distributional preferences is not a central question of this chapter, it allows us to examine how strategic uncertainty affects the relationship between distributional preferences and cooperation.

Table 4.6 presents the results of random effect probit regressions that examine the determinants of first-mover's cooperation in the symmetric games. Column (1) includes Mengel's (2018) payoff indices: RISK ($\frac{P-S}{P}$), TEMPT ($\frac{T-R}{T}$),

and EFF ($\frac{R-P}{R}$). Note that these payoff indices capture important motivations that are relevant in a simultaneous prisoner's dilemmas where subjects face uncertainty about the opponent's choice (Gächter et al., 2020; Mengel, 2018). Similarly, the first-mover faces uncertainty about the second-mover's decision, therefore, we employ the payoff indices of RISK, TEMPT, and EFF. Column (2) adds controls for individual characteristics and session effects in addition to the regressors of column (1). Column (3) adds individual-level parameters of distributional preferences $(\hat{\alpha}, \hat{\rho})$ in addition to the regressors of column (2).

Table 4.6: Determinants of First-movers' Cooperation: Symmetric games

	(1)	(2)	(3)
RISK	-0.010	-0.017	-0.018
	(0.055)	(0.060)	(0.060)
TEMPT	-0.209**	-0.215**	-0.214**
	(0.089)	(0.091)	(0.090)
EFF	0.109**	0.100*	0.099*
	(0.053)	(0.054)	(0.053)
$\hat{\alpha}$			-0.498***
			(0.130)
$\hat{\rho}$			-0.003**
			(0.001)
CCEI			-0.142
			(0.147)
Round	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)
P.Reciprocity		0.027	0.011
		(0.036)	(0.034)
Controls	*No*	*Yes*	*Yes*
Observations	856	832	832
Pseudo R^2	0.024	0.073	0.108
Log-likelihood	-259.2	-238.3	-229.4
BIC	559.0	671.6	673.9

Note: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, elicited using the questionnaire of Falk et al. (2018). Controls: the variables of demographic information, attitudinal characteristics, and session effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

First, TEMPT has a negative impact on first-mover's cooperation. The probability of cooperation decreases by 2.1 - 2.2 percentage points as TEMPT increases by 0.1 point, and the effect is significant at the 5% level. EFF has a positive effect on first-mover's cooperation, but it is marginally significant (i.e. 10% level) once control variables are included (columns (2)-(3)). RISK has an insignificant impact on first-mover's cooperation in all columns (1)-(3).

Second, we find that weaker fair-mindedness hurts first-mover's cooperation. A higher relative weight on subjects' own payoffs (i.e. higher $\hat{\alpha}$) lowers the probability of first-mover's cooperation: based on the complete sample, the probability of cooperation decreases by 5.0 percentage point as $\hat{\alpha}$ increases by 0.1 point (column (3)).

Third, equality-efficiency tradeoffs ($\hat{\rho}$) have a small but significant effect on first-mover's cooperation. Based on the complete sample, subjects who are oriented to equality are more likely to cooperate as a first-mover: the probability of cooperation increases by 0.03 percentage points as $\hat{\rho}$ decreases by 0.1 point.

Lastly, almost all individual characteristics are insignificantly associated with first-mover's cooperation. Among individual characteristics, only the self-assessment of math ability is significant at the 5% level in both columns (2)-(3): subjects who assess themselves as good at math are less likely to cooperate as a first-mover. Positive reciprocity, which is significant for conditional cooperation, is no longer significant for first-mover's cooperation.

Next, we analyze how payoff variations and distributional preferences affect first-mover's cooperation in the asymmetric games (Table 4.7). Columns (1)-(3) present the determinants of first-mover's cooperation in the *Ahead* games, while columns (4)-(6) present the determinants of first-mover's cooperation in

the *Behind* games.

Table 4.7: Determinants of First-movers' Cooperation: Asymmetric games

	Ahead			Behind		
	(1)	(2)	(3)	(4)	(5)	(6)
RISK	0.307	0.245	0.241	0.053	0.065*	0.066*
	(0.277)	(0.286)	(0.282)	(0.036)	(0.037)	(0.037)
TEMPT	-0.362	-0.399	-0.398	-0.192**	-0.196**	-0.199**
	(0.269)	(0.275)	(0.275)	(0.085)	(0.083)	(0.083)
EFF	0.406*	0.343	0.335	0.025	0.030	0.030
	(0.234)	(0.238)	(0.235)	(0.036)	(0.036)	(0.036)
$\hat{\alpha}$			-0.495***			-0.137
			(0.130)			(0.096)
$\hat{\rho}$			-0.003**			-0.001**
			(0.001)			(0.001)
CCEI			0.064			-0.015
			(0.129)			(0.111)
Round	0.000	0.000	0.000	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
P.Reciprocity		0.045	0.029		0.033	0.033*
		(0.036)	(0.033)		(0.022)	(0.019)
Controls	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Observations	856	832	832	856	832	832
Pseudo R^2	0.034	0.081	0.114	0.029	0.130	0.146
Log-likelihood	-275.7	-251.4	-242.6	-174.9	-144.6	-142.0
BIC	591.9	697.9	700.3	390.2	484.1	499.2

Note: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, elicited using the questionnaire of Falk et al. (2018). Controls: the variables of demographic information, attitudinal characteristics, and session effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

First, we observe that the payoff indices are insignificant in the Ahead games while only TEMPT is significant at the 5% level in the Behind games. Second, almost all attitudinal and demographic characteristics are not significantly associated with first-mover's cooperation in the asymmetric games. Third, fair-mindedness is the influential determinant of first-mover's cooperation in the Ahead games: subjects who have a higher relative weight on

subjects' own payoffs (i.e. higher $\hat{\alpha}$) are less likely to cooperate as a first-mover. In the Behind games, fair-mindedness is not significantly associated with the first-mover's cooperation. Lastly, equality-efficiency tradeoffs ($\hat{\rho}$) are significant at the 5% level in both the Ahead and the Behind games, but the effect sizes are very marginal.

Overall, we find that distributional preferences still play a role in cooperation even under strategic uncertainty. Our results support the discussion of Miettinen et al. (2020), which suggests that the first-mover's decision can be accommodated by social preferences and subjective beliefs about the second-mover's cooperation.

4.5.7 Robustness Checks

In Appendix B, we provide several robustness checks. First, we refine the sample by excluding subjects with CCEI below a threshold-level used in previous studies (e.g., 0.80 in Fisman et al.), then run the same econometric specification to the main results. We find that the results using the refined sample are similar to the main results.

Second, we analyze the determinants of free-riding using the same econometric specification to the main results, which presents the determinants of conditional cooperation. We find that stronger fair-mindedness curbs free-riding, which corresponds to the main results that stronger fair-mindedness induces conditional cooperation.

Third, we show that the significant relationship between distributional preferences and conditional cooperation is also valid using other distributional preference models. We estimate individual-level estimates of advantageous inequality aversion (Fehr and Schmidt, 1999) from the modified dictator games,

and show that subjects who reveal a stronger distaste for advantageous inequality are more likely to conditionally cooperate and less likely to free-ride. We also estimate individual-level estimates of the weight on own income and the degree of concern for equality versus maximizing the total income (Charness and Rabin, 2002) from the modified dictator games. We find that subjects who exhibit stronger concern for equality are more likely to conditionally cooperate in the symmetric games, and subjects who have a higher weight on other's income are more likely to conditionally cooperate in the asymmetric games.

4.6 Discussion and Conclusions

We examine the relationship between distributional preferences and conditional cooperation. To do this, we implement a within-subject experiment consisting of two parts: modified dictator games and sequential prisoner's dilemmas. We estimate distributional preferences from modified dictator games, and generate predictions of what each subject will choose as a second-mover in the sequential prisoner's dilemmas. We then compare the predictions with their actual choices to assess the predictive power of distributional preferences.

We find that estimated parameters of subjects' distributional preferences help predict elicited second-mover strategies. More than half of subjects' strategies in twenty-four games are predicted by their distributional preferences at least or greater of 50% accuracy. This is the case using the CES model, but also the case using alternative distributional preference models (i.e. the Fehr-Schmidt and the Charness-Rabin models).

Using the CES model, we divide distributional preferences into fair-mindedness and equality-efficiency tradeoffs. We observe a clear positive relationship be-

tween fair-mindedness and conditional cooperation. Subjects are more likely to conditionally cooperate as they care for other's payoffs as well as their own payoffs. Contrary to this, equality-efficiency tradeoffs are not significantly associated with conditional cooperation. Both equality-oriented and efficiency-oriented subjects are more likely to be conditionally cooperative than self-interested subjects, but we do not find significant differences in conditional cooperation between equality-oriented and efficiency-oriented subjects. We also show that distributional preferences are closely linked to first-movers' cooperation.

Our contribution to existing studies is two fold. First, our experiment employs a broad set of payoff parameterizations in sequential prisoner's dilemmas to study the predictive power of distributional preferences. Most studies that examine the relationship between preferences and behavior in social dilemmas employ a one-shot game with fixed payoff parameters. Second, we show that parsimonious distributional preferences models (Andreoni and Miller, 2002; Charness and Rabin, 2002; Fehr and Schmidt, 1999) can explain how conditional cooperation varies with payoffs.

One question that follows from our study is which experimental method is more suitable for measuring subjects' distributional preferences. Note that studies that did not find a significant relationship between distributional preferences and cooperation in the simultaneous-move social dilemma (Blanco et al., 2011; Haesevoets et al., 2015; Yamagishi et al., 2013) employed a modified dictator game in which the relative price of giving does not vary. In contrast, studies that found the significant role of distributional preferences for cooperation in the simultaneous-move social dilemma experiment (Ghidoni et al. 2019, Hedegaard et al. 2019) employed the discrete choice modified dic-

tator games where the relative price of giving varies. In our study, we vary the relative price of giving in continuous choice modified dictator games: we also find that distributional preferences are significantly associated with the first-mover's cooperation. Therefore, future research could examine whether the mixed results in the literature derive from the differences in elicitation methods.

Our results also have implications for leader selection in the context of social dilemmas. Subjects who substantially consider other's payoffs are more likely to cooperate as a first-mover, while subjects who are close to self-interest mostly defect. This implies that individual distributional preferences also would be the important criteria for selecting a cooperative leader who can induce followers to cooperate through leading-by-example.

References

Afriat, S. "Efficiency Estimation of Production Function." *International Economic Review*, 13(3), 1972, 568–598.

Andreoni, J., and J. Miller. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, 70(2), 2002, 737–753.

Bellemare, C., S. Kröger and A.V. Soest. "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities." *Econometrica*, 76, 2008, 815–839.

Beranek, B., R. Cubitt, and S. Gächter. "Stated and revealed inequality aversion in three subject pools." *Journal of the Economic Science Association*, 1(1), 2015, 43–58.

Blanco, M., D. Engelmann, and H. T. Normann. "A Within-Subject Analysis of Other-Regarding Preferences." *Games and Economic Behavior*, 72, 2011, 321–338.

Bolton, G.E., and A. Ockenfels. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1), 2000, 166–193.

Brandts, J., and G. Charness. "Hot vs. cold: sequential responses in simple experimental games." *Experimental Economics*, 2, 2000, 227–238.

Brandts, J., and G. Charness. "The strategy versus the direct-response method: a first survey of experimental comparisons." *Experimental Economics*, 14, 2011, 375–298.

Bruhin, A., E. Fehr, and D. Schunk. "The many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences." *Journal of the European Economic Association*, 17, 2019, 1025–1069.

Cabrales, A., R. Miniaci, M. Piovesan, and G. Ponti. "Social Preferences and Strategic Uncertainty: An Experiment on Markets and Contracts." *American Economic Review*, 100, 2010, 2261–2278.

Charness, G., and M. Rabin. "Understanding Social Preferences with Simple Tests." *The Quarterly Journal of Economics*, 117(3), 2002, 817–869.

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

Clark, K., and M. Sefton. "The Sequential Prisoner's Dilemma: Evidence on Reciprocation." *The Economic Journal*, 111, 2001, 51–68.

Cooper, D.J., and J.H. Kagel. "Other-regarding preferences." *The handbook of experimental economicsl*, Princeton University Press, 2016.

Costa Lobo, M., and J. Curtice. "Personality Politics?: The Role of Leader Evaluations in Democratic Elections." Oxford University Press, 2015.

Cox, J.C., D. Friedman, and S. Gjerstad. "A Tractable Model of Reciprocity and Fairness." *Games and Economic Behavior*, 59(1), 2007, 17-45.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde. "Global Evidence on Economic Preferences." *Quarterly Journal of Economics*, 133, 2018, 1645–1692.

Fehr, E., and K. M. Schmidt. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114, 1999, 817–868.

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

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

Fisman, R., S. Kariv, and D. Markovits. "Individual Preferences for Giving." *American Economic Review*, 97(5), 2007, 1858–1876.

Fisman, R., P. Jakiela, and S. Kariv. "Distributional Preferences and Political Behavior." *Journal of Public Economics*, 155, 2017, 1–10.

Gächter, S., F. Kölle, and S. Quercia. "Reciprocity and the Tragedies of Maintaining and Providing the Commons." *Nature Human Behavior*, 1, 2017, 650–656.

Gächter, S., K. Lee, and M. Sefton. "Risk, Temptation, and Efficiency in Prisoner's Dilemmas." CeDEx Discussion Paper 2020-15, University of Nottingham, 2020.

Ghidoni, R., B. L. Cleave, and S. Suetens. "Perfect and Imperfect Stragners in Social Dilemmas." *European Economic Review*, 116, 2019, 148–159.

Giamattei, M., K.S. Yahosseini, S. Gachter, and L. Molleman. "LIONESS Lab: a free web-based platform for conducting interactive experiments online." *Journal of Economic Science Association*, forthcoming 2020.

Greiner, B. "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of the Economic Science Association*, 1, 2015, 114–125.

Guan, H., X. Zhu, and P. Zhang. "Rule-Inequality-Aversion Preference and Conditional Cooperation in Public Goods Experiments: Economic Experiment Evidence from China." *Group Decision and Negotiation*,25, 2016, 799–825.

Haesevoets, T., C.R. Folmer, and A. Van Hiel. "Cooperation in in Mixed-Motive Games: The Role of Individual Differences in Selfish and Social Orientation." *European Journal of Personality*, 29, 2015, 445-458.

Hahn, E., J. Gottschling, and F.M. Van Spinath. "Short measurements of personality – Validity and reliability of the GSOEP Big Five Inventory (BFI-S)." *Journal of Research in Personality*, 46, 2012, 355-359.

Hedegaard, M., Kerschbamer, R., Müller, D., Tyran, J.-R. "Distributional preferences explain individual behavior across games and time." SSRN 3388681, 2019.

Katuscak, P., and T. Miklánek. "What Drives Conditional Cooperation in Public Goods Games?" CERGE-EI Working Paper Series No.631, 2018.

Keser, C., and F. van Winden. "Conditional Cooperation and Voluntary Contributions to Public Goods." *Scandinavian Journal of Economics*, 102, 2000, 23-39.

Kroh, M. "Measuring Left-Right Political Orientation: The Choice of Response Format." *Public Opinion Quarterly*, 71(2), 2007, 204–220.

Mengel, F. "Risk and Temptation: A Meta-Study on Prisoner's Dilemma Games." *The Economic Journal*, 128, 2018, 3182–3209.

Miettinen, T., M. Kosfeld, E. Fehr, and J.W. Weibull. "Revealed preferences in a sequential prisoners' dilemma: A horse-race between six utility functions." *Journal of Economic Behavior and Organization*, 173, 2020, 1-25.

Mullett, T., R. McDonald, and G. Brown. "Cooperation in public goods games predicts behavior in incentive-matched binary dilemmas: Evidence for stable prosociality." *Economic Inquiry*, 58(1), 2020, 67–85.

Selten, R. Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes. In: Sauermann, H. (Ed.), Beiträge zur experimentellen Wirtschaftsforschung. J.C.B. Mohr (Paul Siebeck), Tübingen, 1967, 136–168.

Teyssier, S. "Inequity and risk aversion in sequential public good games." *Public Choice*, 151, 2012, 91-119.

Yamagishi, T., N. Mifune, Y. Li, M. Shinada, H. Hashimoto, Y. Horita, A. Miura, K. Inukai, S. Tanida, T. Kiyonari, H. Takagishi, and D. Simunovic. "Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality." *Organizational Behavior and Human Decision Processes*, 120, 2013, 260–271.

Appendices

Appendix 4.A Distributional Preference Models

4.A.1 The Andreoni-Miller Model

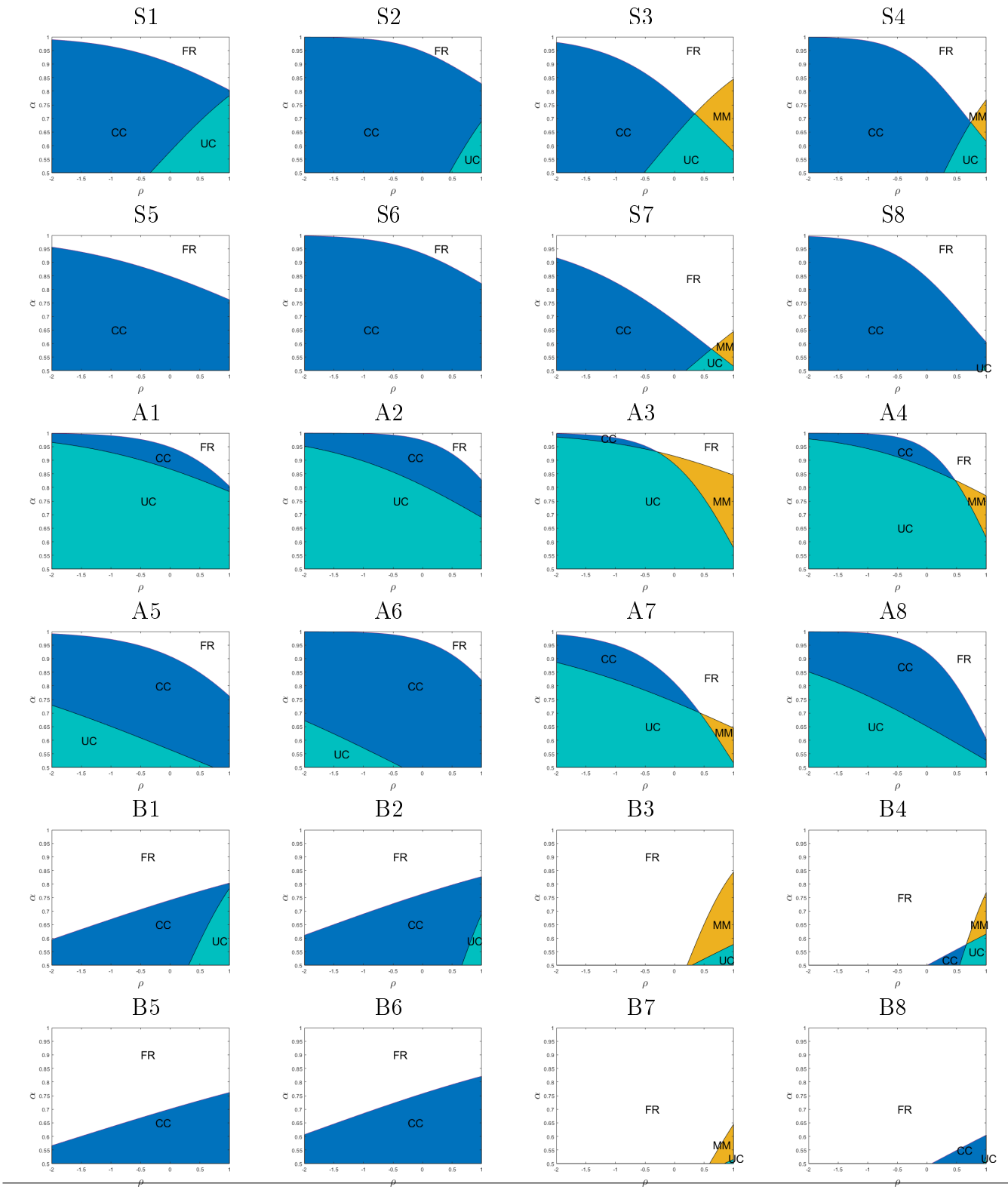
Andreoni and Miller (2002) suggests the CES utility function has the form

$$U_s = [\alpha \pi_s^{\rho} + (1-\alpha)\pi_o^{\rho}]^{1/\rho}$$

where an agent's utility depends on own earnings (π_s) and other's earnings (π_o). α indicates the relative weight on own earnings (the degree of fair-mindedness) and α ranges from 0 to 1, where higher α implies lower degree of fair-mindedness. The parameter ρ, which ranges from -∞ to 1, measures the curvature of indifference curves in (π_s, π_o) space.

Figure 4.A1 illustrates the preference space of (α, ρ) for second-mover strategies - conditional cooperation, free-riding, unconditional cooperation, and mis-matching - across twenty-four games. The likelihood of second-mover strategies is predicted to differ across games. In the asymmetric games with advantageous payoffs (i.e. A1 to A8), the likelihood of unconditional cooperation is predicted to notably increase compared to other games. In the asymmetric games with disadvantageous payoffs (i.e. B1 to B8), the likelihood of free-riding is predicted to remarably increase than other games.

Figure 4.A1: Predicted Second-mover's Strategies of the CES model



Note: S - Symmetric games, A - Asymmetric games with advantageous payoffs (Ahead), B - Asymmetric games with disadvantageous payoffs (Behind).

4.A.2 The Fehr-Schmidt Model

The Fehr-Schmidt (henceforth FS) model is based on the assumption that agents have preferences that can be represented by utility functions of the following form.

$$U_s = \begin{cases} \pi_s - \beta \max\{\pi_s - \pi_o, 0\} & \text{if } \pi_s > \pi_o \\ \pi_s - \gamma \max\{\pi_o - \pi_s, 0\} & \text{if } \pi_s < \pi_o \end{cases}$$

, where $\beta \leq \gamma$ and $0 \leq \beta < 1$. β indicates the distaste of advantageous inequality, and γ indicates the distaste of disadvantageous inequality. When $\gamma = \beta = 0$, the utility is only affected by one's own payoff and does not have any relevance to other's payoff: therefore, such subject can be classified as purely self-interested. In case of $\gamma \geq \beta > 0$, a subject exhibits difference aversion in that he/she put more weights on his/her own payoffs when his/her payoff is smaller than other's and he put the positive weight on other's payoff when his/her payoff is greater than other's.

Econometric Specification

From the FS utility function, the marginal rate of substitution and the relative price of self and other's earnings given by as follows.

$$\frac{MU_o}{MU_s} = \frac{p_o}{p_s} \quad \Rightarrow \quad \frac{\beta}{1-\beta} = \frac{p_o}{p_s}$$

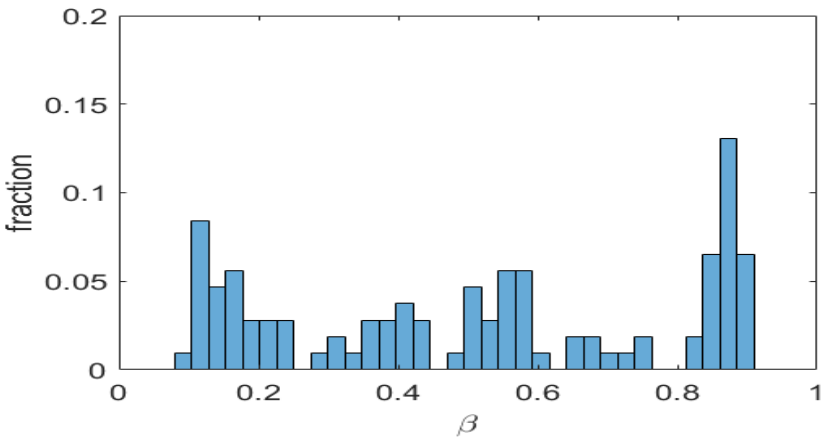
Due to the piecewise linearity of the FS utility function, the model predicts that if the relative price of giving ($\frac{p_o}{p_s}$) is greater than $\frac{\beta_s}{1-\beta_s}$, the agent will allocate the entire budget for self ($\pi_s = \frac{m}{p_s}, \pi_o = 0$), and if $\frac{p_o}{p_s}$ is smaller than $\frac{\beta_s}{1-\beta_s}$, the agent will allocate the same amount of earnings to him/herself and

the other person ($\pi_s = \pi_o = \frac{m}{p_s+p_o}$). Lastly, if $\frac{p_o}{p_s}$ equals to $\frac{\beta_s}{1-\beta_s}$, any allocation for him/herself $\pi_s \in [\frac{m}{p_s+p_o}, \frac{m}{p_s}]$ can be chosen.

Our estimate of β for an individual is the value that minimizes the root mean squared error (RMSE) between actual and predicted allocations to self across the 50 modified dictator games. Note that γ, the disadavantageous inequality parameter cannot be estimated from the data.

Figure 4.A2 provides the overview of subjects' individual advantageous inequality aversion. Among 107 subjects, 28 subjects show weak advantageous inequality-aversion ($\hat{\beta} < 0.235$), 21 subjects show moderate advantageous inequality-aversion ($0.235 \leq \hat{\beta} < 0.5$), and 58 subjects show strong advantageous inequality-aversion ($0.5 \leq \hat{\beta}$).[18]

Figure 4.A2: Individual-level Estimates of $\hat{\beta}$



Predictions of Conditional Cooperation

The FS model predicts that i) the second-mover cooperates in response to first-mover's cooperation if and only if the utility of mutual cooperation

[18] For the threshold of weak (0.235), moderate (0.235 $\sim$ 0.5), and strong (0.5 $\sim$) advantageous inequality aversion, we follow the classification of Fehr and Schmidt (1999). We also compared the distribution of β with Beranek et al. (2015)'s Nottingham samples, and we do not find significant differences.

outweighs that of defecting against cooperator (i.e. $R > [(1-\beta)T + \beta S]$), and ii) the second-mover never cooperates in response to first-mover's defection due to non-negativity of γ (i.e. $S - \gamma(T-S) < P$).

Therefore, The second-mover's utility maximizing strategy depends on preferences β and payoffs (T, R, P, S) as follows. In a symmetric sequential prisoner's dilemma, the second-mover is predicted to free-ride if $\beta < \frac{T-R}{T-S}$ and coniditonally cooperate if $\beta > \frac{T-R}{T-S}$.

Figure 4.A3 illustrates the preference space of (β) for second-mover strategies in twenty-four SPDs. The Nash equilibrium assuming self-interested agent ($\beta = 0$) is unilateral defection regardless of payoff variations. For non-selfish agents ($\beta > 0$), the second-mover strategy depends on the degree of advantageous inequality aversion and payoffs.

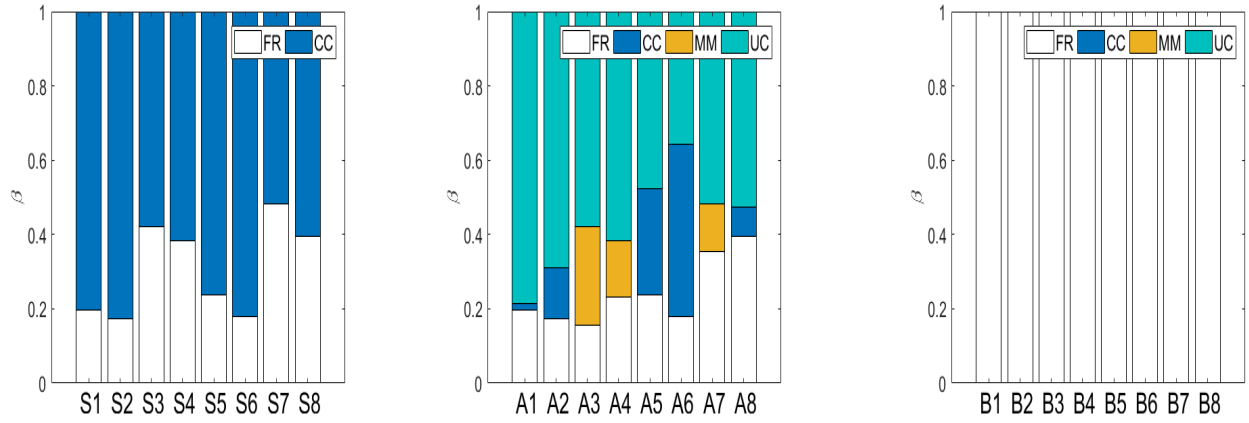
First, the likelihood of second-mover strategies is predicted to differ across games where we vary payoffs and payoff structure (i.e. symmetric and asymmetric payoffs). In the symmetric games (i.e. S1 to S8), the likelihood of conditional cooperation is predicted to increase with DAMAGE (e.g., S1 with lower DAMAGE and S2 with higher DAMAGE), but to decrease with GAIN (e.g. S1 with lower GAIN and S3 with higher GAIN). Payoff asymmetry affects the likelihood of second-mover strategy as follows. In the asymmetric games with advantageous payoffs (i.e. A1 to A8), the likelihood of unconditional cooperation is predicted to notably increase compared to other games. In the asymmetric games with disadvantageous payoffs (i.e. B1 to B8), free-riding is predicted.

Second, β has a clear effect on the likelihood of conditional cooperation and free-riding in the symmetric games: higher β induces conditional cooperation but deters free-riding. In the asymmetric games with advantageous payoffs,

β has a clear positive effect on unconditional cooperation. In the asymmetric games with disadvantageous payoffs, β does not have any effect on the second-mover strategy because all subjects are predicted to free-ride regardless of β.

Figure 4.A3: Predicted Second-mover's Strategies of the FS model

(a) Symmetric Games (b) Asymmetric Games: Ahead (c) Asymmetric Games: Behind



Note: S - Symmetric games, A - Asymmetric games with advantageous payoffs (Ahead), B - Asymmetric games with disadvantageous payoffs (Behind).

4.A.3 The Charness-Rabin Model

The Charness-Rabin (henceforth CR) model is based on the assumption that agents have preferences that can be represented by utility functions of the following form.

$$U_s = \lambda\pi_s + (1-\lambda)[\delta min(\pi_s, \pi_o) + (1-\delta)(\pi_s + \pi_o)]$$

, where $\lambda \in [0,1]$ and $\delta \in [0,1]$. λ indicates the weight on own income, and δ indicates the degree of concern for equality versus maximizing the total income. When $\lambda = 1$, the utility is only affected by one's own payoff and does not have any relevance to other's payoff: therefore, such subject can be

classified as purely self-interested. When $\delta = 1$, it corresponds to Maximin preferences, while $\delta = 0$ corresponds to maximizing total surplus.

Econometric Specification

From the CR utility function, the marginal rate of substitution and the relative price of self and other's earnings given by as follows.

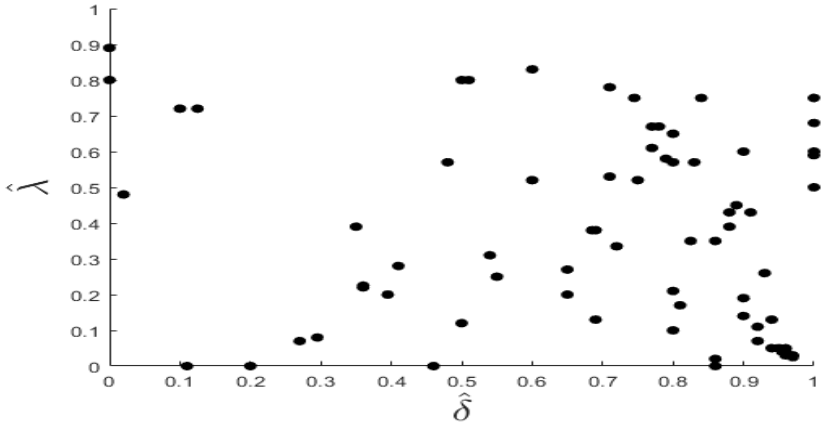
$$\text{if } \pi_s > \pi_o, \quad \frac{MU_o}{MU_s} = \frac{p_o}{p_s} \quad \Rightarrow \quad \frac{1-\lambda}{1-\delta(1-\lambda)} = \frac{p_o}{p_s}$$

$$\text{if } \pi_s < \pi_o, \quad \frac{MU_o}{MU_s} = \frac{p_o}{p_s} \quad \Rightarrow \quad \frac{(1-\lambda)(1-\delta)}{1} = \frac{p_o}{p_s}$$

Due to the piecewise linearity of the CR utility function, the model predicts either allocating the same amount between him/herself and the other person, allocating the entire budget for self, or allocating the entire budget for others. If the relative price of giving ($\frac{p_o}{p_s}$) is smaller than $\frac{1-\lambda}{1-\delta(1-\lambda)}$ but greater than $\frac{(1-\lambda)(1-\delta)}{1}$, the agent will allocate the same amount of earnings to him/herself and the other person ($\pi_s = \pi_o = \frac{m}{p_s+p_o}$). If the relative price of giving ($\frac{p_o}{p_s}$) is smaller than $\frac{1-\lambda}{1-\delta(1-\lambda)}$ and $\frac{(1-\lambda)(1-\delta)}{1}$, the agent will either allocate the same amount of earnings to him/herself and the other person or give all earnigs to the other person (depending on the relative size of utilities). If the relative price of giving ($\frac{p_o}{p_s}$) is greater than $\frac{1-\lambda}{1-\delta(1-\lambda)}$ and $\frac{(1-\lambda)(1-\delta)}{1}$, the agent will either allocate the same amount of earnings to him/herself and the other person or take all earnigs for self (depending on the relative size of utilities). Lastly, if the relative price of giving ($\frac{p_o}{p_s}$) is greater than $\frac{1-\lambda}{1-\delta(1-\lambda)}$ but smaller than $\frac{(1-\lambda)(1-\delta)}{1}$, the agent will either take all earnings for self or give all earnigs to the other person (depending on the relative size of utilities).

Our estimate of λ and δ for an individual is the value that minimizes the root mean squared error (RMSE) between actual and predicted allocations to self across the 50 modified dictator games. Figure 4.A4 provides the overview of subjects' individual-level estimates of ($\hat{\lambda}$, $\hat{\delta}$). Among 107 subjects, 22 subjects (20.6%) exhibit a higher relative weight on earnings for self ($\hat{\lambda} \geq 0.8$). 8 subjects (7.5%) exhibit a strong concern for efficiency ($\hat{\lambda} < 0.8$ and $\hat{\delta} < 0.33$), 13 subjects (12.1%) exhibit a moderate concern for efficiency and equality ($\hat{\lambda} < 0.8$ and $0.33 \leq \hat{\delta} < 0.66$), and 64 subjects (59.8%) exhibit a strong concern for equality ($\hat{\lambda} < 0.8$ and $0.66 \leq \hat{\delta}$).[19]

Figure 4.A4: Individual-level Estimates of $\hat{\lambda}$ and $\hat{\delta}$



[19] We compare individual-level estimates of ($\hat{\alpha}$, $\hat{\rho}$) of the CES model with ($\hat{\lambda}$, $\hat{\delta}$) of the Charness-Rabin model. $\hat{\alpha}$ and $\hat{\lambda}$ are strongly correlated (Spearman correlation: 0.761), which implies that subjects who have a higher weight on earnings for self in the CES model are also likely to have a higher weight on earnings for self in the CR model. $\hat{\rho}$ and $\hat{\delta}$ have a negative correlation (-0.310), which suggests that subjects who are estimated as equality-oriented in the CES model (lower $\hat{\rho}$ are likely to be estimated as equality-oriented in the CR model (higher $\hat{\delta}$).

Appendix 4.B Robustness Checks

Tables 4.B1-B12 provide the regression results using the distributional preference parameters of the CES model ($\hat{\alpha}$, $\hat{\rho}$): Tables 4.B1-B2 provide the results for conditional cooperation, Tables 4.B3-B4 provide the results for free-riding, and Tables 4.B5-B6 provide the results for first-mover's cooperation. Tables 4.B7-12 provide corresponding results for the subsample of subjects with CCEI > 0.80.

Tables 4.B13-B24 provide the regression results using the advantageous inequality aversion parameter of the Fehr-Schmidt model ($\hat{\beta}$): Tables 4.B13-B14 provide the results for conditional cooperation, Tables 4.B15-B16 provide the results for free-riding, and Tables B17-B18 provide the results for first-mover's cooperation. Tables 4.B19-24 provide corresponding results for the subsample of subjects with CCEI > 0.80.

Tables 4.B25-B36 provide the regression results using the distributional preference parameters of the Charness-Rabin model ($\hat{\lambda}$, $\hat{\delta}$): Tables 4.B25-B26 provide the results for conditional cooperation, Tables 4.B27-B28 provide the results for free-riding, and Tables 4.B29-B30 provide the results for first-mover's cooperation. Tables 4.B31-36 provide corresponding results for the subsample of subjects with CCEI > 0.80.

Table 4.B1: Determinants of Conditional Cooperation: Symmetric games

	All	All	All
DAMAGE	0.216***	0.208***	0.206***
	(0.075)	(0.076)	(0.075)
GAIN	-0.140**	-0.137**	-0.136**
	(0.058)	(0.059)	(0.058)
EFF	-0.008	-0.021	-0.022
	(0.066)	(0.066)	(0.066)
$\hat{\alpha}$			-0.849***
			(0.132)
$\hat{\rho}$			-0.002
			(0.002)
CCEI			-0.068
			(0.178)
Round	0.001	0.001	0.001
	(0.002)	(0.001)	(0.001)
	the measures of Falk et al. (2018)		
P.Reciprocity		0.136***	0.105***
		(0.045)	(0.039)
N.Reciprocity		-0.014	-0.020
		(0.049)	(0.042)
Patience		0.027	0.037
		(0.047)	(0.042)
Risk taking		0.065	0.038
		(0.043)	(0.040)
Trust		0.030	0.012
		(0.032)	(0.030)
Math		-0.034	-0.054*
		(0.036)	(0.030)
Procrastination		-0.032	0.008
		(0.047)	(0.042)
	Big-Five personality using Hahn et al. (2012)		
FiveC		-0.021	0.032
		(0.040)	(0.040)
FiveA		-0.012	-0.007
		(0.047)	(0.038)
FiveE		-0.040	-0.026
		(0.028)	(0.026)
FiveN		0.008	-0.020
		(0.032)	(0.030)
FiveO		0.010	0.002
		(0.043)	(0.034)

	All	All	All
		Demographic characteristics	
Politics		-0.054***	-0.042**
		(0.019)	(0.018)
Economics		-0.275**	-0.210*
		(0.109)	(0.107)
Age		-0.006	-0.005
		(0.006)	(0.006)
Female		-0.047	-0.042
		(0.067)	(0.058)
		Expenditure per month, excluding rent - baseline: £ 400 to 800	
$<$ £ 400		-0.174*	-0.106
		(0.090)	(0.080)
$>$ £ 800		-0.283***	-0.266***
		(0.105)	(0.086)
		Ethnicity - baseline: Others	
Asian		-0.171*	-0.058
		(0.088)	(0.072)
White		-0.099	-0.137
		(0.096)	(0.087)
		Session - baseline: Session 1	
Session 2		0.040	0.118
		(0.085)	(0.082)
Session 3		0.141	0.150*
		(0.089)	(0.089)
Session 4		-0.087	-0.025
		(0.096)	(0.086)
Control	*No*	*Yes*	*Yes*
Obs.	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B2: Determinants of Conditional Cooperation: Asymmetric games

	Ahead	Ahead	Ahead	Behind	Behind	Behind
DAMAGE	0.099	0.074	0.075	0.895***	0.829***	0.832***
	(0.089)	(0.085)	(0.085)	(0.313)	(0.297)	(0.288)
GAIN	-0.563***	-0.594***	-0.585***	-0.151*	-0.158**	-0.150*
	(0.167)	(0.168)	(0.165)	(0.082)	(0.080)	(0.078)
EFF	-0.390***	-0.397***	-0.390***	-0.025	-0.019	-0.018
	(0.149)	(0.150)	(0.147)	(0.053)	(0.053)	(0.051)
$\hat{\alpha}$			-0.300***			-0.497***
			(0.082)			(0.075)
$\hat{\rho}$			0.001			-0.001**
			(0.001)			(0.001)
CCEI			0.106			0.017
			(0.121)			(0.094)
Round	-0.002	-0.003	-0.003	-0.003	-0.003*	-0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			the measures of Falk et al. (2018)			
P.Reciprocity		0.059***	0.043**		0.038	0.019
		(0.022)	(0.020)		(0.028)	(0.021)
N.Reciprocity		-0.012	-0.015		-0.020	-0.022
		(0.023)	(0.022)		(0.024)	(0.018)
Patience		-0.003	-0.001		-0.003	-0.002
		(0.024)	(0.021)		(0.026)	(0.021)
Risk taking		0.029	0.014		0.062**	0.040
		(0.022)	(0.021)		(0.030)	(0.025)
Trust		0.007	0.008		0.024	0.007
		(0.017)	(0.018)		(0.018)	(0.015)
Math		-0.003	-0.005		0.020	0.005
		(0.018)	(0.016)		(0.016)	(0.011)
Procrastination		-0.013	0.004		-0.018	-0.001
		(0.023)	(0.022)		(0.029)	(0.022)
			Big-Five personality using Hahn et al. (2012)			
FiveC		0.002	0.014		-0.009	0.013
		(0.022)	(0.021)		(0.024)	(0.021)
FiveA		-0.008	-0.001		0.007	0.006
		(0.024)	(0.022)		(0.024)	(0.017)
FiveE		-0.007	0.007		-0.034**	-0.020*
		(0.013)	(0.014)		(0.017)	(0.012)
FiveN		0.004	-0.009		0.035**	0.015
		(0.013)	(0.013)		(0.017)	(0.013)
FiveO		0.014	0.012		0.005	0.002
		(0.021)	(0.019)		(0.017)	(0.013)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
			Demographic characteristics			
Politics		-0.026***	-0.022**		-0.019*	-0.010
		(0.010)	(0.010)		(0.010)	(0.009)
Economics		-0.140**	-0.138**		-0.109	-0.074
		(0.064)	(0.059)		(0.069)	(0.056)
Age		0.001	0.002		0.000	0.001
		(0.003)	(0.003)		(0.004)	(0.003)
Female		-0.011	0.010		-0.079*	-0.066*
		(0.037)	(0.038)		(0.042)	(0.037)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		-0.123***	-0.084**		0.001	0.039
		(0.046)	(0.042)		(0.051)	(0.035)
$>$ £ 800		-0.147***	-0.143***		-0.015	-0.031
		(0.055)	(0.043)		(0.062)	(0.049)
			Ethnicity - baseline: Others			
Asian		-0.084*	-0.054		-0.103**	-0.022
		(0.045)	(0.045)		(0.047)	(0.043)
White		-0.073	-0.099**		-0.059	-0.064
		(0.047)	(0.046)		(0.060)	(0.048)
			Session - baseline: Session 1			
Session 2		0.060	0.089**		0.015	0.065
		(0.041)	(0.042)		(0.051)	(0.043)
Session 3		0.072	0.070		0.064	0.064
		(0.052)	(0.051)		(0.050)	(0.044)
Session 4		-0.049	-0.020		0.042	0.082
		(0.053)	(0.051)		(0.064)	(0.057)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	856	832	832	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B3: Determinants of Free Riding: Symmetric games

	All	All	All
DAMAGE	-0.166*	-0.160*	-0.161*
	(0.087)	(0.088)	(0.089)
GAIN	0.167***	0.166***	0.169***
	(0.057)	(0.057)	(0.058)
EFF	-0.047	-0.034	-0.034
	(0.057)	(0.057)	(0.058)
$\hat{\alpha}$			0.848***
			(0.153)
$\hat{\rho}$			0.002
			(0.002)
CCEI			0.302
			(0.219)
Round	-0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		-0.147***	-0.118***
		(0.048)	(0.043)
N.Reciprocity		0.019	0.039
		(0.055)	(0.048)
Patience		-0.051	-0.066
		(0.050)	(0.045)
Risk taking		-0.055	-0.025
		(0.044)	(0.039)
Trust		-0.028	0.005
		(0.033)	(0.031)
Math		0.022	0.048
		(0.035)	(0.029)
Procrastination		0.058	0.010
		(0.049)	(0.042)
	Big-Five personality using Hahn et al. (2012)		
FiveC		0.070	0.006
		(0.047)	(0.048)
FiveA		-0.024	-0.023
		(0.051)	(0.043)
FiveE		0.041	0.031
		(0.031)	(0.029)
FiveN		-0.006	0.018
		(0.031)	(0.028)
FiveO		-0.034	-0.028
		(0.044)	(0.036)

	All	All	All
		Demographic characteristics	
Politics		0.031	0.021
		(0.021)	(0.018)
Economics		0.262**	0.151
		(0.117)	(0.109)
Age		0.010	0.010*
		(0.007)	(0.006)
Female		0.047	0.065
		(0.073)	(0.064)
		Expenditure per month, excluding rent - baseline: £ 400 to 800	
$<$ £ 400		0.166*	0.109
		(0.091)	(0.080)
$>$ £ 800		0.117	0.105
		(0.132)	(0.110)
		Ethnicity - baseline: Others	
Asian		0.170**	0.033
		(0.086)	(0.074)
White		0.084	0.107
		(0.091)	(0.088)
		Session - baseline: Session 1	
Session 2		-0.041	-0.097
		(0.091)	(0.088)
Session 3		-0.143	-0.132
		(0.098)	(0.097)
Session 4		0.072	0.027
		(0.094)	(0.088)
Control	*No*	*Yes*	*Yes*
Obs.	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B4: Determinants of Free Riding: Asymmetric games

	Ahead	Ahead	Ahead	Behind	Behind	Behind
DAMAGE	-0.134	-0.111	-0.112	-0.545*	-0.531*	-0.525*
	(0.096)	(0.097)	(0.097)	(0.316)	(0.319)	(0.310)
GAIN	0.262*	0.278*	0.280*	0.162*	0.160*	0.151*
	(0.150)	(0.154)	(0.155)	(0.086)	(0.088)	(0.085)
EFF	-0.293**	-0.265**	-0.265**	-0.075	-0.067	-0.066
	(0.131)	(0.132)	(0.132)	(0.047)	(0.048)	(0.047)
$\hat{\alpha}$			0.775***			0.679***
			(0.161)			(0.100)
$\hat{\rho}$			0.003*			0.001
			(0.002)			(0.001)
CCEI			-0.112			0.196
			(0.206)			(0.151)
Round	-0.001	-0.002	-0.001	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			the measures of Falk et al. (2018)			
P.Reciprocity		-0.127***	-0.098**		-0.057	-0.020
		(0.046)	(0.040)		(0.040)	(0.030)
N.Reciprocity		0.022	0.018		0.022	0.035
		(0.052)	(0.045)		(0.035)	(0.028)
Patience		-0.037	-0.046		-0.013	-0.024
		(0.047)	(0.041)		(0.036)	(0.029)
Risk taking		-0.068	-0.045		-0.031	0.005
		(0.043)	(0.040)		(0.042)	(0.036)
Trust		0.008	0.010		-0.025	0.003
		(0.032)	(0.029)		(0.026)	(0.020)
Math		0.012	0.022		-0.034	-0.017
		(0.034)	(0.031)		(0.027)	(0.019)
Procrastination		0.046	0.013		0.029	-0.003
		(0.046)	(0.040)		(0.042)	(0.033)
			Big-Five personality using Hahn et al. (2012)			
FiveC		0.069	0.024		0.066*	0.028
		(0.045)	(0.047)		(0.036)	(0.033)
FiveA		-0.044	-0.052		-0.036	-0.033
		(0.051)	(0.043)		(0.032)	(0.024)
FiveE		0.010	-0.006		0.032	0.018
		(0.032)	(0.031)		(0.024)	(0.018)
FiveN		-0.011	0.016		-0.026	-0.003
		(0.029)	(0.026)		(0.025)	(0.020)
FiveO		-0.026	-0.018		-0.030	-0.026
		(0.039)	(0.034)		(0.025)	(0.018)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
			Demographic characteristics			
Politics		0.060***	0.045**		0.000	-0.012
		(0.020)	(0.018)		(0.016)	(0.013)
Economics		0.081	0.036		0.138	0.065
		(0.113)	(0.099)		(0.107)	(0.087)
Age		0.011	0.010*		0.004	0.003
		(0.006)	(0.005)		(0.005)	(0.004)
Female		0.051	0.014		0.090	0.098*
		(0.073)	(0.071)		(0.063)	(0.054)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		0.106	0.053		-0.041	-0.090*
		(0.086)	(0.069)		(0.065)	(0.049)
$>$ £ 800		0.166	0.131		-0.107	-0.098
		(0.129)	(0.116)		(0.115)	(0.091)
			Ethnicity - baseline: Others			
Asian		0.149*	0.058		0.164**	0.041
		(0.089)	(0.080)		(0.071)	(0.063)
White		0.123	0.157*		0.091	0.087
		(0.093)	(0.084)		(0.083)	(0.064)
			Session - baseline: Session 1			
Session 2		-0.058	-0.141		-0.065	-0.107*
		(0.089)	(0.087)		(0.070)	(0.056)
Session 3		-0.077	-0.089		-0.045	-0.021
		(0.095)	(0.095)		(0.077)	(0.069)
Session 4		0.104	0.031		-0.083	-0.111
		(0.098)	(0.093)		(0.087)	(0.078)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	856	832	832	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B5: Determinants of First-mover's Cooperation: Symmetric games

	All	All	All
RISK	-0.010	-0.017	-0.018
	(0.055)	(0.060)	(0.060)
TEMPT	-0.209**	-0.215**	-0.214**
	(0.089)	(0.091)	(0.090)
EFF	0.109**	0.100*	0.099*
	(0.053)	(0.054)	(0.053)
$\hat{\alpha}$			-0.498***
			(0.130)
$\hat{\rho}$			-0.003**
			(0.001)
CCEI			-0.142
			(0.147)
Round	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		0.027	0.011
		(0.036)	(0.034)
N.Reciprocity		-0.021	-0.019
		(0.043)	(0.042)
Patience		-0.007	0.001
		(0.032)	(0.028)
Risk taking		0.073**	0.060*
		(0.034)	(0.032)
Trust		0.029	0.019
		(0.028)	(0.028)
Math		-0.074**	-0.096***
		(0.035)	(0.033)
Procrastination		-0.007	0.003
		(0.046)	(0.042)
	Big-Five personality using Hahn et al. (2012)		
FiveC		-0.043	-0.012
		(0.037)	(0.036)
FiveA		0.023	0.025
		(0.042)	(0.037)
FiveE		-0.050*	-0.056**
		(0.028)	(0.026)
FiveN		-0.012	-0.033
		(0.027)	(0.024)
FiveO		0.018	0.009
		(0.033)	(0.035)

	All	All	All
		Demographic characteristics	
Politics		-0.021	-0.005
		(0.017)	(0.016)
Economics		-0.055	-0.018
		(0.089)	(0.094)
Age		-0.007	-0.008
		(0.006)	(0.006)
Female		0.021	0.006
		(0.066)	(0.063)
		Expenditure per month, excluding rent - baseline: £ 400 to 800	
$<$ £ 400		0.058	0.088
		(0.065)	(0.060)
$>$ £ 800		0.030	0.047
		(0.074)	(0.070)
		Ethnicity - baseline: Others	
Asian		0.026	0.103
		(0.077)	(0.074)
White		0.001	0.014
		(0.078)	(0.069)
		Session - baseline: Session 1	
Session 2		0.023	0.080
		(0.070)	(0.074)
Session 3		0.029	0.032
		(0.084)	(0.088)
Session 4		-0.060	-0.038
		(0.085)	(0.082)
Control	*No*	*Yes*	*Yes*
Obs.	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B6: Determinants of First-mover's Cooperation: Asymmetric games

	Ahead	Ahead	Ahead	Behind	Behind	Behind
RISK	0.307	0.245	0.241	0.053	0.065*	0.066*
	(0.277)	(0.286)	(0.282)	(0.036)	(0.037)	(0.037)
TEMPT	-0.362	-0.399	-0.398	-0.192**	-0.196**	-0.199**
	(0.269)	(0.275)	(0.275)	(0.085)	(0.083)	(0.083)
EFF	0.406*	0.343	0.335	0.025	0.030	0.030
	(0.234)	(0.238)	(0.235)	(0.036)	(0.036)	(0.036)
$\hat{\alpha}$			-0.495***			-0.137
			(0.130)			(0.096)
$\hat{\rho}$			-0.003**			-0.001**
			(0.001)			(0.001)
CCEI			0.064			-0.015
			(0.129)			(0.111)
Round	0.000	0.000	0.000	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
			the measures of Falk et al. (2018)			
P.Reciprocity		0.045	0.029		0.033	0.033*
		(0.036)	(0.033)		(0.022)	(0.019)
N.Reciprocity		0.036	0.040		0.006	0.013
		(0.039)	(0.037)		(0.024)	(0.023)
Patience		0.004	0.009		0.013	0.014
		(0.035)	(0.033)		(0.019)	(0.017)
Risk taking		0.037	0.019		0.018	0.015
		(0.035)	(0.032)		(0.018)	(0.018)
Trust		0.014	0.006		0.009	0.010
		(0.021)	(0.023)		(0.014)	(0.015)
Math		-0.014	-0.026		-0.043**	-0.050***
		(0.025)	(0.022)		(0.018)	(0.018)
Procrastination		-0.044	-0.041		-0.036	-0.040
		(0.041)	(0.036)		(0.025)	(0.025)
			Big-Five personality using Hahn et al. (2012)			
FiveC		-0.056	-0.033		-0.017	-0.007
		(0.036)	(0.036)		(0.019)	(0.019)
FiveA		0.033	0.039		0.002	-0.003
		(0.033)	(0.029)		(0.023)	(0.021)
FiveE		-0.049**	-0.041*		-0.016	-0.022*
		(0.024)	(0.022)		(0.012)	(0.012)
FiveN		0.006	-0.006		0.023*	0.017
		(0.023)	(0.020)		(0.012)	(0.013)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
FiveO		0.014	0.009		0.031*	0.030*
		(0.027)	(0.028)		(0.018)	(0.018)
			Demographic characteristics			
Politics		-0.011	0.001		-0.004	0.002
		(0.014)	(0.013)		(0.010)	(0.010)
Economics		-0.184*	-0.168		-0.093	-0.095
		(0.104)	(0.104)		(0.072)	(0.071)
Age		-0.013**	-0.015**		-0.005	-0.006**
		(0.006)	(0.006)		(0.003)	(0.003)
Female		0.063	0.071		-0.049	-0.050
		(0.055)	(0.053)		(0.032)	(0.030)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		0.034	0.071		0.081***	0.079**
		(0.059)	(0.050)		(0.029)	(0.033)
$>$ £ 800		0.097	0.135		-0.002	-0.004
		(0.111)	(0.115)		(0.026)	(0.028)
			Ethnicity - baseline: Others			
Asian		-0.098	-0.036		-0.022	-0.025
		(0.079)	(0.076)		(0.039)	(0.040)
White		-0.117	-0.134*		0.006	-0.009
		(0.079)	(0.069)		(0.045)	(0.045)
			Session - baseline: Session 1			
Session 2		0.046	0.106*		-0.019	0.000
		(0.064)	(0.064)		(0.041)	(0.042)
Session 3		0.078	0.092		-0.012	0.000
		(0.071)	(0.070)		(0.043)	(0.046)
Session 4		0.072	0.120*		-0.076	-0.069
		(0.075)	(0.070)		(0.049)	(0.044)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	856	832	832	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B7: Determinants of Conditional Cooperation: Symmetric games & CCEI > 0.80

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
DAMAGE	0.168**	0.161**	0.158**
	(0.077)	(0.079)	(0.077)
GAIN	-0.188***	-0.191***	-0.189***
	(0.064)	(0.065)	(0.065)
EFF	0.004	-0.013	-0.012
	(0.073)	(0.076)	(0.074)
$\hat{\alpha}$			-0.794***
			(0.160)
$\hat{\rho}$			0.000
			(0.001)
CCEI			-1.092*
			(0.623)
Round	0.001	0.000	0.000
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		0.115**	0.118***
		(0.047)	(0.043)
N.Reciprocity		-0.029	-0.035
		(0.053)	(0.043)
Patience		0.051	0.044
		(0.049)	(0.043)
Risk taking		0.048	0.039
		(0.048)	(0.040)
Trust		0.041	0.033
		(0.040)	(0.032)
Math		-0.039	-0.047
		(0.046)	(0.038)
Procrastination		-0.056	-0.013
		(0.053)	(0.046)
	Big-Five personality using Hahn et al. (2012)		
FiveC		-0.024	0.023
		(0.041)	(0.041)
FiveA		-0.027	-0.010
		(0.048)	(0.041)
FiveE		-0.071**	-0.056**
		(0.031)	(0.026)
FiveN		0.003	-0.015
		(0.032)	(0.028)

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
FiveO		0.009	-0.018
		(0.049)	(0.039)
	Demographic characteristics		
Politics		-0.068***	-0.042*
		(0.024)	(0.023)
Economics		-0.180	-0.171*
		(0.115)	(0.103)
Age		-0.011*	-0.005
		(0.006)	(0.007)
Female		-0.141*	-0.081
		(0.081)	(0.079)
	Expenditure per month, excluding rent - baseline: £ 400 to 800		
< £ 400		-0.268***	-0.191**
		(0.099)	(0.083)
> £ 800		-0.255*	-0.302***
		(0.131)	(0.088)
	Ethnicity - baseline: Others		
Asian		-0.128	-0.102
		(0.100)	(0.080)
White		0.028	-0.058
		(0.117)	(0.100)
	Session - baseline: Session 1		
Session 2		0.036	0.089
		(0.092)	(0.084)
Session 3		0.067	0.060
		(0.098)	(0.096)
Session 4		-0.091	-0.028
		(0.107)	(0.088)
Control	*No*	*Yes*	*Yes*
Obs.	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B8: Determinants of Conditional Cooperation: Asymmetric games & CCEI > 0.80

	Ahead	Ahead	Ahead	Behind	Behind	Behind
DAMAGE	0.168*	0.143*	0.144	0.939***	0.851***	0.853***
	(0.093)	(0.087)	(0.088)	(0.325)	(0.300)	(0.286)
GAIN	-0.574***	-0.619***	-0.611***	-0.193**	-0.205**	-0.195**
	(0.192)	(0.194)	(0.195)	(0.089)	(0.086)	(0.083)
EFF	-0.444**	-0.449***	-0.433**	-0.048	-0.039	-0.042
	(0.175)	(0.174)	(0.175)	(0.052)	(0.050)	(0.048)
$\hat{\alpha}$			-0.317***			-0.593***
			(0.085)			(0.105)
$\hat{\rho}$			0.001*			-0.001
			(0.001)			(0.001)
CCEI			-0.062			-0.242
			(0.344)			(0.266)
Round	-0.002	-0.003	-0.002	-0.002	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			the measures of Falk et al. (2018)			
P.Reciprocity		0.045**	0.042*		0.040	0.040
		(0.023)	(0.022)		(0.035)	(0.027)
N.Reciprocity		-0.038	-0.038		-0.059**	-0.063***
		(0.025)	(0.024)		(0.030)	(0.022)
Patience		0.008	0.002		-0.001	-0.012
		(0.025)	(0.022)		(0.031)	(0.024)
Risk taking		0.012	0.004		0.036	0.019
		(0.024)	(0.022)		(0.033)	(0.024)
Trust		0.016	0.018		0.032	0.024
		(0.020)	(0.018)		(0.021)	(0.015)
Math		0.019	0.014		0.018	0.012
		(0.022)	(0.019)		(0.020)	(0.014)
Procrastination		-0.014	0.001		-0.007	0.004
		(0.029)	(0.026)		(0.035)	(0.025)
			Big-Five personality using Hahn et al. (2012)			
FiveC		-0.012	-0.001		0.014	0.022
		(0.023)	(0.022)		(0.027)	(0.022)
FiveA		-0.018	-0.005		-0.027	-0.009
		(0.026)	(0.023)		(0.026)	(0.019)
FiveE		-0.023	-0.011		-0.030	-0.015
		(0.015)	(0.015)		(0.020)	(0.012)
FiveN		0.016	0.010		0.040*	0.021
		(0.016)	(0.014)		(0.021)	(0.017)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
FiveO		0.039*	0.036*		0.003	-0.006
		(0.023)	(0.021)		(0.022)	(0.019)
			Demographic characteristics			
Politics		-0.020	-0.007		-0.022*	0.007
		(0.013)	(0.013)		(0.013)	(0.012)
Economics		-0.117*	-0.134**		-0.088	-0.097*
		(0.061)	(0.059)		(0.072)	(0.055)
Age		-0.000	0.003		-0.002	0.002
		(0.004)	(0.004)		(0.005)	(0.004)
Female		-0.034	0.002		-0.101*	-0.045
		(0.045)	(0.044)		(0.053)	(0.044)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		-0.158***	-0.119**		-0.028	0.025
		(0.056)	(0.047)		(0.064)	(0.045)
$>$ £ 800		-0.112	-0.140***		0.017	-0.046
		(0.073)	(0.051)		(0.074)	(0.054)
			Ethnicity - baseline: Others			
Asian		-0.096*	-0.085*		-0.060	-0.046
		(0.050)	(0.049)		(0.062)	(0.053)
White		-0.063	-0.093*		0.005	-0.057
		(0.055)	(0.050)		(0.076)	(0.057)
			Session - baseline: Session 1			
Session 2		0.072	0.091**		0.012	0.092**
		(0.047)	(0.045)		(0.058)	(0.046)
Session 3		0.099	0.085		0.064	0.073
		(0.061)	(0.057)		(0.057)	(0.051)
Session 4		-0.009	0.013		0.031	0.114*
		(0.054)	(0.047)		(0.075)	(0.062)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	688	664	664	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B9: Determinants of Free Riding: Symmetric games & CCEI > 0.80

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
DAMAGE	-0.181**	-0.174**	-0.175**
	(0.087)	(0.088)	(0.088)
GAIN	0.184***	0.184***	0.186***
	(0.063)	(0.064)	(0.065)
EFF	-0.068	-0.053	-0.054
	(0.065)	(0.066)	(0.067)
$\hat{\alpha}$			0.833***
			(0.168)
$\hat{\rho}$			-0.001
			(0.001)
CCEI			1.233*
			(0.695)
Round	-0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		-0.124**	-0.132***
		(0.051)	(0.048)
N.Reciprocity		0.048	0.062
		(0.059)	(0.047)
Patience		-0.066	-0.060
		(0.050)	(0.047)
Risk taking		-0.048	-0.040
		(0.049)	(0.040)
Trust		-0.031	-0.020
		(0.041)	(0.033)
Math		0.046	0.060
		(0.045)	(0.038)
Procrastination		0.055	0.010
		(0.054)	(0.047)
	Big-Five personality using Hahn et al. (2012)		
FiveC		0.046	-0.008
		(0.046)	(0.048)
FiveA		0.016	0.001
		(0.051)	(0.045)
FiveE		0.098***	0.081***
		(0.035)	(0.031)
FiveN		-0.017	-0.007
		(0.031)	(0.028)
FiveO		-0.037	-0.011
		(0.050)	(0.040)

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
		Demographic characteristics	
Politics		0.058**	0.029
		(0.024)	(0.024)
Economics		0.102	0.078
		(0.133)	(0.111)
Age		0.014**	0.009
		(0.006)	(0.007)
Female		0.155*	0.102
		(0.084)	(0.087)
		Expenditure per month, excluding rent - baseline: £ 400 to 800	
< £ 400		0.262***	0.182**
		(0.098)	(0.085)
> £ 800		0.136	0.179*
		(0.136)	(0.096)
		Ethnicity - baseline: Others	
Asian		0.138	0.101
		(0.103)	(0.083)
White		-0.033	0.054
		(0.114)	(0.101)
		Session - baseline: Session 1	
Session 2		-0.026	-0.067
		(0.096)	(0.088)
Session 3		-0.140	-0.125
		(0.107)	(0.103)
Session 4		0.063	-0.002
		(0.112)	(0.094)
Control	*No*	*Yes*	*Yes*
Obs.	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B10: Determinants of Free Riding: Asymmetric games & CCEI > 0.80

	Ahead	Ahead	Ahead	Behind	Behind	Behind
DAMAGE	-0.138	-0.111	-0.115	-0.610*	-0.600*	-0.600*
	(0.110)	(0.112)	(0.113)	(0.339)	(0.345)	(0.335)
GAIN	0.187	0.205	0.208	0.185*	0.183*	0.175*
	(0.170)	(0.177)	(0.180)	(0.098)	(0.099)	(0.097)
EFF	-0.442***	-0.414***	-0.419***	-0.103**	-0.096**	-0.091**
	(0.142)	(0.142)	(0.142)	(0.045)	(0.045)	(0.043)
$\hat{\alpha}$			0.864***			0.729***
			(0.156)			(0.132)
$\hat{\rho}$			0.001			0.001
			(0.002)			(0.001)
CCEI			0.492			0.455
			(0.605)			(0.354)
Round	-0.002	-0.003	-0.002	-0.002	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			the measures of Falk et al. (2018)			
P.Reciprocity		-0.097**	-0.092**		-0.058	-0.053
		(0.047)	(0.041)		(0.044)	(0.034)
N.Reciprocity		0.031	0.032		0.075*	0.084***
		(0.052)	(0.043)		(0.041)	(0.031)
Patience		-0.043	-0.031		-0.002	0.007
		(0.047)	(0.038)		(0.036)	(0.026)
Risk taking		-0.075	-0.061		-0.009	0.007
		(0.047)	(0.039)		(0.041)	(0.032)
Trust		-0.009	-0.006		-0.021	-0.015
		(0.037)	(0.027)		(0.026)	(0.019)
Math		-0.010	0.002		-0.004	0.004
		(0.042)	(0.033)		(0.027)	(0.020)
Procrastination		0.066	0.019		-0.007	-0.024
		(0.053)	(0.045)		(0.047)	(0.035)
			Big-Five personality using Hahn et al. (2012)			
FiveC		0.082*	0.028		0.001	-0.014
		(0.044)	(0.048)		(0.036)	(0.031)
FiveA		-0.027	-0.045		0.008	-0.006
		(0.054)	(0.042)		(0.034)	(0.027)
FiveE		0.061*	0.039		0.035	0.016
		(0.034)	(0.031)		(0.027)	(0.019)
FiveN		-0.023	0.002		-0.038	-0.019
		(0.030)	(0.025)		(0.027)	(0.021)
FiveO		-0.053	-0.031		-0.021	-0.007
		(0.048)	(0.039)		(0.027)	(0.025)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
			Demographic characteristics			
Politics		0.068***	0.038		0.018	-0.016
		(0.023)	(0.024)		(0.016)	(0.016)
Economics		0.024	0.018		0.039	0.039
		(0.117)	(0.095)		(0.098)	(0.072)
Age		0.013*	0.008		0.007	0.003
		(0.006)	(0.006)		(0.006)	(0.005)
Female		0.117	0.040		0.139**	0.081
		(0.085)	(0.088)		(0.064)	(0.054)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		0.214**	0.143*		-0.030	-0.092*
		(0.098)	(0.075)		(0.072)	(0.056)
$>$ £ 800		0.200	0.245**		-0.071	-0.003
		(0.142)	(0.097)		(0.090)	(0.061)
			Ethnicity - baseline: Others			
Asian		0.193*	0.158*		0.086	0.065
		(0.107)	(0.086)		(0.082)	(0.071)
White		0.089	0.183**		0.023	0.094
		(0.114)	(0.084)		(0.096)	(0.072)
			Session - baseline: Session 1			
Session 2		-0.104	-0.182**		-0.058	-0.151**
		(0.102)	(0.089)		(0.078)	(0.061)
Session 3		-0.170*	-0.162*		-0.065	-0.068
		(0.098)	(0.091)		(0.081)	(0.071)
Session 4		0.016	-0.058		-0.063	-0.153*
		(0.104)	(0.080)		(0.095)	(0.079)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	688	664	664	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B11: Determinants of First-mover's Cooperation: Symmetric games & CCEI > 0.80

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
RISK	-0.053	-0.065	-0.066
	(0.061)	(0.065)	(0.065)
TEMPT	-0.225**	-0.229**	-0.226**
	(0.100)	(0.097)	(0.095)
EFF	0.143**	0.128**	0.129**
	(0.057)	(0.053)	(0.052)
$\hat{\alpha}$			-0.373**
			(0.161)
$\hat{\rho}$			-0.001
			(0.001)
CCEI			-0.863*
			(0.481)
Round	-0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		-0.007	-0.004
		(0.034)	(0.033)
N.Reciprocity		-0.043	-0.056
		(0.041)	(0.042)
Patience		-0.002	0.008
		(0.031)	(0.032)
Risk taking		0.064*	0.065**
		(0.036)	(0.032)
Trust	0.030	0.026	
		(0.031)	(0.031)
Math		-0.095**	-0.105***
		(0.039)	(0.037)
Procrastination		-0.014	-0.018
		(0.042)	(0.038)
	Big-Five personality using Hahn et al. (2012)		
FiveC		-0.039	-0.036
		(0.033)	(0.029)
FiveA		-0.002	0.004
		(0.040)	(0.036)
FiveE		-0.086***	-0.087***
		(0.028)	(0.023)
FiveN		0.003	-0.008
		(0.024)	(0.024)

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
FiveO		0.042	0.022
		(0.033)	(0.037)
	Demographic characteristics		
Politics		-0.043**	-0.028
		(0.018)	(0.020)
Economics		0.016	0.036
		(0.090)	(0.088)
Age		-0.000	0.000
		(0.005)	(0.005)
Female		-0.080	-0.087
		(0.066)	(0.065)
	Expenditure per month, excluding rent - baseline: £ 400 to 800		
< £ 400		0.002	0.040
		(0.076)	(0.074)
> £ 800		0.031	0.003
		(0.076)	(0.065)
	Ethnicity - baseline: Others		
Asian		0.174***	0.205***
		(0.061)	(0.060)
White		0.137**	0.118*
		(0.069)	(0.061)
	Session - baseline: Session 1		
Session 2		-0.013	0.019
		(0.072)	(0.076)
Session 3		0.144*	0.133
		(0.079)	(0.081)
Session 4		-0.124	-0.104
		(0.094)	(0.084)
Control	*No*	*Yes*	*Yes*
Obs.	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B12: Determinants of First-mover's Cooperation: Asymmetric games & CCEI > 0.80

	Ahead	Ahead	Ahead	Behind	Behind	Behind
RISK	0.395	0.343	0.331	0.061	0.075*	0.077*
	(0.274)	(0.291)	(0.283)	(0.040)	(0.040)	(0.039)
TEMPT	-0.274	-0.332	-0.337	-0.224**	-0.227**	-0.229**
	(0.276)	(0.296)	(0.294)	(0.101)	(0.093)	(0.093)
EFF	0.526**	0.484*	0.473**	0.037	0.045	0.045
	(0.245)	(0.251)	(0.240)	(0.041)	(0.040)	(0.040)
$\hat{\alpha}$			-0.538***			-0.090
			(0.139)			(0.083)
$\hat{\rho}$			0.000			-0.001**
			(0.001)			(0.001)
CCEI			0.119			0.212
			(0.402)			(0.347)
Round	-0.001	-0.001	-0.001	-0.002	-0.002*	-0.002*
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
			the measures of Falk et al. (2018)			
P.Reciprocity		0.044	0.026		0.027	0.022
		(0.035)	(0.032)		(0.018)	(0.017)
N.Reciprocity		-0.007	-0.011		-0.020	-0.007
		(0.040)	(0.038)		(0.018)	(0.020)
Patience		0.005	-0.002		0.012	0.012
		(0.034)	(0.030)		(0.016)	(0.015)
Risk taking		0.035	0.017		-0.008	-0.010
		(0.037)	(0.031)		(0.013)	(0.016)
Trust		0.009	0.011		0.009	0.015
		(0.024)	(0.021)		(0.012)	(0.013)
Math		-0.011	-0.019		-0.031*	-0.038**
		(0.028)	(0.026)		(0.018)	(0.019)
Procrastination		-0.015	-0.004		-0.040	-0.049*
		(0.048)	(0.040)		(0.025)	(0.026)
			Big-Five personality using Hahn et al. (2012)			
FiveC		-0.054	-0.035		-0.030*	-0.025*
		(0.036)	(0.035)		(0.017)	(0.015)
FiveA		0.024	0.043		-0.016	-0.015
		(0.031)	(0.028)		(0.018)	(0.017)
FiveE		-0.073***	-0.054**		-0.014	-0.018**
		(0.025)	(0.022)		(0.009)	(0.009)
FiveN		0.031	0.026		0.014	0.007
		(0.022)	(0.020)		(0.010)	(0.011)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
FiveO		0.033	0.035		0.035**	0.034**
		(0.029)	(0.032)		(0.015)	(0.014)
			Demographic characteristics			
Politics		-0.010	0.012		-0.011	-0.007
		(0.017)	(0.016)		(0.010)	(0.011)
Economics		-0.049	-0.087		-0.033	-0.035
		(0.109)	(0.124)		(0.044)	(0.046)
Age		-0.010	-0.009*		-0.005*	-0.006**
		(0.007)	(0.005)		(0.003)	(0.003)
Female		0.076	0.098*		-0.081***	-0.087***
		(0.064)	(0.055)		(0.029)	(0.033)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		0.006	0.047		0.006	0.011
		(0.062)	(0.053)		(0.033)	(0.041)
$>$ £ 800		0.164	0.113		-0.019	-0.018
		(0.138)	(0.123)		(0.036)	(0.035)
			Ethnicity - baseline: Others			
Asian		-0.034	-0.019		-0.008	-0.030
		(0.081)	(0.073)		(0.033)	(0.042)
White		-0.043	-0.118*		0.028	-0.015
		(0.085)	(0.069)		(0.040)	(0.040)
			Session - baseline: Session 1			
Session 2		0.086	0.135**		0.008	0.038
		(0.067)	(0.065)		(0.033)	(0.033)
Session 3		0.191**	0.185**		0.038	0.056
		(0.081)	(0.074)		(0.039)	(0.042)
Session 4		0.131	0.186**		-0.072	-0.042
		(0.087)	(0.082)		(0.044)	(0.034)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	688	664	664	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B13: Determinants of Conditional Cooperation in the Fehr-Schmidt model: Symmetric games

	All	All	All
DAMAGE	0.216***	0.208***	0.208***
	(0.075)	(0.076)	(0.075)
GAIN	-0.140**	-0.137**	-0.137**
	(0.058)	(0.059)	(0.058)
EFF	-0.008	-0.021	-0.022
	(0.066)	(0.066)	(0.066)
$\hat{\beta}$			0.610***
			(0.112)
CCEI			0.018
			(0.179)
Round	0.001	0.001	0.001
	(0.002)	(0.001)	(0.001)
	the measures of Falk et al. (2018)		
P.Reciprocity		0.136***	0.099**
		(0.045)	(0.039)
N.Reciprocity		-0.014	0.001
		(0.049)	(0.043)
Patience		0.027	0.051
		(0.047)	(0.047)
Risk taking		0.065	0.061
		(0.043)	(0.039)
Trust		0.030	-0.006
		(0.032)	(0.033)
Math		-0.034	-0.028
		(0.036)	(0.034)
Procrastination		-0.032	-0.007
		(0.047)	(0.046)
	Big-Five personality using Hahn et al. (2012)		
FiveC		-0.021	0.005
		(0.040)	(0.038)
FiveA		-0.012	0.008
		(0.047)	(0.040)
FiveE		-0.040	-0.067***
		(0.028)	(0.025)
FiveN		0.008	-0.025
		(0.032)	(0.032)
FiveO		0.010	0.000
		(0.043)	(0.039)

	All	All	All
		Demographic characteristics	
Politics		-0.054***	-0.070***
		(0.019)	(0.017)
Economics		-0.275**	-0.193*
		(0.109)	(0.110)
Age		-0.006	-0.009
		(0.006)	(0.006)
Female		-0.047	-0.101*
		(0.067)	(0.057)
		Expenditure per month, excluding rent - baseline: £ 400 to 800	
$<$ £ 400		-0.174*	-0.178**
		(0.090)	(0.080)
$>$ £ 800		-0.283***	-0.270***
		(0.105)	(0.103)
		Ethnicity - baseline: Others	
Asian		-0.171*	-0.092
		(0.088)	(0.069)
White		-0.099	-0.103
		(0.096)	(0.092)
		Session - baseline: Session 1	
Session 2		0.040	0.107
		(0.085)	(0.083)
Session 3		0.141	0.185**
		(0.089)	(0.094)
Session 4		-0.087	-0.013
		(0.096)	(0.090)
Control	*No*	*Yes*	*Yes*
Obs.	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B14: Determinants of Conditional Cooperation in the Fehr-Schmidt model: Asymmetric games

	Ahead	Ahead	Ahead	Behind	Behind	Behind
DAMAGE	0.099	0.074	0.070	0.895***	0.829***	0.854***
	(0.089)	(0.085)	(0.084)	(0.313)	(0.297)	(0.297)
GAIN	-0.563***	-0.594***	-0.589***	-0.151*	-0.158**	-0.153*
	(0.167)	(0.168)	(0.165)	(0.082)	(0.080)	(0.080)
EFF	-0.390***	-0.397***	-0.386***	-0.025	-0.019	-0.016
	(0.149)	(0.150)	(0.146)	(0.053)	(0.053)	(0.053)
$\hat{\beta}$			0.236***			0.404***
			(0.062)			(0.070)
CCEI			0.160			0.052
			(0.130)			(0.101)
Round	-0.002	-0.003	-0.003*	-0.003	-0.003*	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			the measures of Falk et al. (2018)			
P.Reciprocity		0.059***	0.040*		0.038	0.014
		(0.022)	(0.021)		(0.028)	(0.023)
N.Reciprocity		-0.012	-0.002		-0.020	-0.014
		(0.023)	(0.022)		(0.024)	(0.020)
Patience		-0.003	0.002		-0.003	0.008
		(0.024)	(0.022)		(0.026)	(0.021)
Risk taking		0.029	0.028		0.062**	0.054**
		(0.022)	(0.020)		(0.030)	(0.024)
Trust		0.007	0.003		0.024	-0.006
		(0.017)	(0.019)		(0.018)	(0.017)
Math		-0.003	0.004		0.020	0.024*
		(0.018)	(0.017)		(0.016)	(0.013)
Procrastination		-0.013	-0.006		-0.018	-0.014
		(0.023)	(0.023)		(0.029)	(0.023)
			Big-Five personality using Hahn et al. (2012)			
FiveC		0.002	0.007		-0.009	-0.002
		(0.022)	(0.020)		(0.024)	(0.021)
FiveA		-0.008	0.006		0.007	0.010
		(0.024)	(0.022)		(0.024)	(0.020)
FiveE		-0.007	-0.011		-0.034**	-0.044***
		(0.013)	(0.013)		(0.017)	(0.015)
FiveN		0.004	-0.009		0.035**	0.016
		(0.013)	(0.014)		(0.017)	(0.015)
FiveO		0.014	0.011		0.005	-0.005
		(0.021)	(0.019)		(0.017)	(0.014)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
			Demographic characteristics			
Politics		-0.026***	-0.029***		-0.019*	-0.027***
		(0.010)	(0.010)		(0.010)	(0.010)
Economics		-0.140**	-0.130**		-0.109	-0.059
		(0.064)	(0.057)		(0.069)	(0.061)
Age		0.001	0.000		0.000	-0.002
		(0.003)	(0.003)		(0.004)	(0.003)
Female		-0.011	-0.015		-0.079*	-0.105***
		(0.037)	(0.036)		(0.042)	(0.036)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		-0.123***	-0.120***		0.001	0.015
		(0.046)	(0.046)		(0.051)	(0.041)
$>$ £ 800		-0.147***	-0.143***		-0.015	0.006
		(0.055)	(0.054)		(0.062)	(0.051)
			Ethnicity - baseline: Others			
Asian		-0.084*	-0.066		-0.103**	-0.050
		(0.045)	(0.043)		(0.047)	(0.043)
White		-0.073	-0.088*		-0.059	-0.056
		(0.047)	(0.045)		(0.060)	(0.052)
			Session - baseline: Session 1			
Session 2		0.060	0.100**		0.015	0.068
		(0.041)	(0.042)		(0.051)	(0.048)
Session 3		0.072	0.100*		0.064	0.087
		(0.052)	(0.055)		(0.050)	(0.053)
Session 4		-0.049	0.001		0.042	0.096*
		(0.053)	(0.050)		(0.064)	(0.055)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	856	832	832	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B15: Determinants of Free Riding in the Fehr-Schmidt model: Symmetric games

	All	All	All
DAMAGE	-0.166*	-0.160*	-0.160*
	(0.087)	(0.088)	(0.088)
GAIN	0.167***	0.166***	0.167***
	(0.057)	(0.057)	(0.057)
EFF	-0.047	-0.034	-0.034
	(0.057)	(0.057)	(0.058)
$\hat{\beta}$			-0.671***
			(0.111)
CCEI			0.143
			(0.206)
Round	-0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		-0.147***	-0.103**
		(0.048)	(0.042)
N.Reciprocity		0.019	0.011
		(0.055)	(0.047)
Patience		-0.051	-0.074
		(0.050)	(0.049)
Risk taking		-0.055	-0.056
		(0.044)	(0.037)
Trust		-0.028	0.026
		(0.033)	(0.033)
Math		0.022	0.021
		(0.035)	(0.030)
Procrastination		0.058	0.034
		(0.049)	(0.044)
	Big-Five personality using Hahn et al. (2012)		
FiveC		0.070	0.035
		(0.047)	(0.043)
FiveA		-0.024	-0.044
		(0.051)	(0.043)
FiveE		0.041	0.073***
		(0.031)	(0.026)
FiveN		-0.006	0.023
		(0.031)	(0.029)
FiveO		-0.034	-0.028
		(0.044)	(0.040)

	All	All	All
		Demographic characteristics	
Politics		0.031	0.048***
		(0.021)	(0.017)
Economics		0.262**	0.140
		(0.117)	(0.113)
Age		0.010	0.014**
		(0.007)	(0.006)
Female		0.047	0.134**
		(0.073)	(0.060)
		Expenditure per month, excluding rent - baseline: £ 400 to 800	
$<$ £ 400		0.166*	0.179**
		(0.091)	(0.076)
$>$ £ 800		0.117	0.113
		(0.132)	(0.116)
		Ethnicity - baseline: Others	
Asian		0.170**	0.066
		(0.086)	(0.068)
White		0.084	0.084
		(0.091)	(0.092)
		Session - baseline: Session 1	
Session 2		-0.041	-0.107
		(0.091)	(0.086)
Session 3		-0.143	-0.185*
		(0.098)	(0.100)
Session 4		0.072	-0.010
		(0.094)	(0.085)
Control	*No*	*Yes*	*Yes*
Obs.	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B16: Determinants of Free Riding in the Fehr-Schmidt model: Asymmetric games

	Ahead	Ahead	Ahead	Behind	Behind	Behind
DAMAGE	-0.134	-0.111	-0.111	-0.545*	-0.531*	-0.537*
	(0.096)	(0.097)	(0.097)	(0.316)	(0.319)	(0.315)
GAIN	0.262*	0.278*	0.282*	0.162*	0.160*	0.155*
	(0.150)	(0.154)	(0.155)	(0.086)	(0.088)	(0.086)
EFF	-0.293**	-0.265**	-0.266**	-0.075	-0.067	-0.068
	(0.131)	(0.132)	(0.133)	(0.047)	(0.048)	(0.048)
$\hat{\beta}$			-0.629***			-0.549***
			(0.101)			(0.095)
CCEI			-0.241			0.093
			(0.199)			(0.150)
Round	-0.001	-0.002	-0.001	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			the measures of Falk et al. (2018)			
P.Reciprocity		-0.127***	-0.076*		-0.057	-0.011
		(0.046)	(0.039)		(0.040)	(0.034)
N.Reciprocity		0.022	-0.002		0.022	0.018
		(0.052)	(0.044)		(0.035)	(0.030)
Patience		-0.037	-0.054		-0.013	-0.029
		(0.047)	(0.044)		(0.036)	(0.031)
Risk taking		-0.068	-0.067*		-0.031	-0.016
		(0.043)	(0.035)		(0.042)	(0.033)
Trust		0.008	0.026		-0.025	0.021
		(0.032)	(0.031)		(0.026)	(0.022)
Math		0.012	0.001		-0.034	-0.040*
		(0.034)	(0.029)		(0.027)	(0.021)
Procrastination		0.046	0.025		0.029	0.019
		(0.046)	(0.042)		(0.042)	(0.034)
			Big-Five personality using Hahn et al. (2012)			
FiveC		0.069	0.048		0.066*	0.046
		(0.045)	(0.042)		(0.036)	(0.030)
FiveA		-0.044	-0.074*		-0.036	-0.041
		(0.051)	(0.042)		(0.032)	(0.026)
FiveE		0.010	0.028		0.032	0.049**
		(0.032)	(0.030)		(0.024)	(0.019)
FiveN		-0.011	0.024		-0.026	-0.006
		(0.029)	(0.027)		(0.025)	(0.022)
FiveO		-0.026	-0.013		-0.030	-0.018
		(0.039)	(0.035)		(0.025)	(0.021)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
			Demographic characteristics			
Politics		0.060***	0.072***		0.000	0.009
		(0.020)	(0.017)		(0.016)	(0.014)
Economics		0.081	0.012		0.138	0.043
		(0.113)	(0.104)		(0.107)	(0.099)
Age		0.011	0.013**		0.004	0.008*
		(0.006)	(0.006)		(0.005)	(0.005)
Female		0.051	0.089		0.090	0.154***
		(0.073)	(0.064)		(0.063)	(0.052)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		0.106	0.107		-0.041	-0.057
		(0.086)	(0.071)		(0.065)	(0.053)
$>$ £ 800		0.166	0.147		-0.107	-0.144
		(0.129)	(0.121)		(0.115)	(0.088)
			Ethnicity - baseline: Others			
Asian		0.149*	0.074		0.164**	0.078
		(0.089)	(0.078)		(0.071)	(0.066)
White		0.123	0.139*		0.091	0.082
		(0.093)	(0.084)		(0.083)	(0.073)
			Session - baseline: Session 1			
Session 2		-0.058	-0.150*		-0.065	-0.124**
		(0.089)	(0.086)		(0.070)	(0.062)
Session 3		-0.077	-0.137		-0.045	-0.064
		(0.095)	(0.095)		(0.077)	(0.081)
Session 4		0.104	-0.008		-0.083	-0.139*
		(0.098)	(0.084)		(0.087)	(0.076)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	856	832	832	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B17: Determinants of First-mover's Cooperation in the Fehr-Schmidt model: Symmetric games

	All	All	All
RISK	-0.010	-0.017	-0.019
	(0.055)	(0.060)	(0.061)
TEMPT	-0.209**	-0.215**	-0.218**
	(0.089)	(0.091)	(0.091)
EFF	0.109**	0.100*	0.101*
	(0.053)	(0.054)	(0.054)
$\hat{\beta}$			0.413***
			(0.098)
CCEI			-0.111
			(0.163)
Round	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		0.027	-0.011
		(0.036)	(0.036)
N.Reciprocity		-0.021	-0.006
		(0.043)	(0.039)
Patience		-0.007	0.016
		(0.032)	(0.031)
Risk taking		0.073**	0.062**
		(0.034)	(0.030)
Trust		0.029	0.003
		(0.028)	(0.030)
Math		-0.074**	-0.075**
		(0.035)	(0.032)
Procrastination		-0.007	0.005
		(0.046)	(0.042)
	Big-Five personality using Hahn et al. (2012)		
FiveC		-0.043	-0.027
		(0.037)	(0.034)
FiveA		0.023	0.042
		(0.042)	(0.040)
FiveE		-0.050*	-0.070***
		(0.028)	(0.027)
FiveN		-0.012	-0.038
		(0.027)	(0.026)
FiveO		0.018	0.005
		(0.033)	(0.032)

	All	All	All
		Demographic characteristics	
Politics		-0.021	-0.028*
		(0.017)	(0.015)
Economics		-0.055	0.017
		(0.089)	(0.101)
Age		-0.007	-0.010
		(0.006)	(0.007)
Female		0.021	-0.045
		(0.066)	(0.063)
		Expenditure per month, excluding rent - baseline: £ 400 to 800	
$<$ £ 400		0.058	0.047
		(0.065)	(0.059)
$>$ £ 800		0.030	0.044
		(0.074)	(0.073)
		Ethnicity - baseline: Others	
Asian		0.026	0.093
		(0.077)	(0.067)
White		0.001	0.019
		(0.078)	(0.075)
		Session - baseline: Session 1	
Session 2		0.023	0.069
		(0.070)	(0.069)
Session 3		0.029	0.053
		(0.084)	(0.092)
Session 4		-0.060	-0.014
		(0.085)	(0.079)
Control	*No*	*Yes*	*Yes*
Obs.	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B18: Determinants of First-mover's Cooperation in the Fehr-Schmidt model: Asymmetric games

	Ahead	Ahead	Ahead	Behind	Behind	Behind
RISK	0.307	0.245	0.246	0.053	0.065*	0.066*
	(0.277)	(0.286)	(0.285)	(0.036)	(0.037)	(0.037)
TEMPT	-0.362	-0.399	-0.403	-0.192**	-0.196**	-0.199**
	(0.269)	(0.275)	(0.279)	(0.085)	(0.083)	(0.083)
EFF	0.406*	0.343	0.340	0.025	0.030	0.030
	(0.234)	(0.238)	(0.237)	(0.036)	(0.036)	(0.036)
$\hat{\beta}$			0.613***			0.115
			(0.107)			(0.078)
CCEI			0.165			-0.040
			(0.144)			(0.124)
Round	0.000	0.000	-0.000	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
			the measures of Falk et al. (2018)			
P.Reciprocity		0.045	-0.001		0.033	0.025
		(0.036)	(0.031)		(0.022)	(0.021)
N.Reciprocity		0.036	0.045		0.006	0.007
		(0.039)	(0.029)		(0.024)	(0.022)
Patience		0.004	0.038		0.013	0.023
		(0.035)	(0.035)		(0.019)	(0.019)
Risk taking		0.037	0.032		0.018	0.019
		(0.035)	(0.026)		(0.018)	(0.018)
Trust		0.014	-0.032		0.009	-0.000
		(0.021)	(0.024)		(0.014)	(0.015)
Math		-0.014	-0.011		-0.043**	-0.042***
		(0.025)	(0.020)		(0.018)	(0.016)
Procrastination		-0.044	-0.039		-0.036	-0.033
		(0.041)	(0.031)		(0.025)	(0.023)
			Big-Five personality using Hahn et al. (2012)			
FiveC		-0.056	-0.048*		-0.017	-0.011
		(0.036)	(0.027)		(0.019)	(0.018)
FiveA		0.033	0.049**		0.002	0.002
		(0.033)	(0.025)		(0.023)	(0.022)
FiveE		-0.049**	-0.067***		-0.016	-0.024**
		(0.024)	(0.021)		(0.012)	(0.012)
FiveN		0.006	-0.015		0.023*	0.017
		(0.023)	(0.021)		(0.012)	(0.012)
FiveO		0.014	0.000		0.031*	0.028
		(0.027)	(0.024)		(0.018)	(0.017)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
			Demographic characteristics			
Politics		-0.011	-0.020*		-0.004	-0.006
		(0.014)	(0.012)		(0.010)	(0.009)
Economics		-0.184*	-0.148		-0.093	-0.085
		(0.104)	(0.095)		(0.072)	(0.074)
Age		-0.013**	-0.018***		-0.005	-0.007**
		(0.006)	(0.005)		(0.003)	(0.003)
Female		0.063	0.009		-0.049	-0.061*
		(0.055)	(0.046)		(0.032)	(0.032)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		0.034	0.071		0.081***	0.075**
		(0.059)	(0.045)		(0.029)	(0.030)
$>$ £ 800		0.097	0.171*		-0.002	0.000
		(0.111)	(0.098)		(0.026)	(0.031)
			Ethnicity - baseline: Others			
Asian		-0.098	-0.038		-0.022	-0.009
		(0.079)	(0.063)		(0.039)	(0.038)
White		-0.117	-0.130**		0.006	0.006
		(0.079)	(0.064)		(0.045)	(0.041)
			Session - baseline: Session 1			
Session 2		0.046	0.124**		-0.019	-0.021
		(0.064)	(0.059)		(0.041)	(0.040)
Session 3		0.078	0.104		-0.012	-0.015
		(0.071)	(0.070)		(0.043)	(0.047)
Session 4		0.072	0.170***		-0.076	-0.071*
		(0.075)	(0.054)		(0.049)	(0.042)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	856	832	832	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B19: Determinants of Conditional Cooperation in the Fehr-Schmidt model: Symmetric games & CCEI > 0.80

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
DAMAGE	0.168**	0.161**	0.159**
	(0.077)	(0.079)	(0.078)
GAIN	-0.188***	-0.191***	-0.188***
	(0.064)	(0.065)	(0.065)
EFF	0.004	-0.013	-0.013
	(0.073)	(0.076)	(0.074)
$\hat{\beta}$			0.499***
			(0.120)
CCEI			-0.808
			(0.643)
Round	0.001	0.000	0.000
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		0.115**	0.114***
		(0.047)	(0.041)
N.Reciprocity		-0.029	-0.019
		(0.053)	(0.045)
Patience		0.051	0.053
		(0.049)	(0.047)
Risk taking		0.048	0.056
		(0.048)	(0.040)
Trust		0.041	0.012
		(0.040)	(0.036)
Math		-0.039	-0.033
		(0.046)	(0.041)
Procrastination		-0.056	-0.042
		(0.053)	(0.046)
	Big-Five personality using Hahn et al. (2012)		
FiveC		-0.024	0.004
		(0.041)	(0.038)
FiveA		-0.027	-0.010
		(0.048)	(0.042)
FiveE		-0.071**	-0.076***
		(0.031)	(0.025)
FiveN		0.003	-0.020
		(0.032)	(0.029)
FiveO		0.009	-0.015
		(0.049)	(0.046)

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
		Demographic characteristics	
Politics		-0.068***	-0.072***
		(0.024)	(0.021)
Economics		-0.180	-0.152
		(0.115)	(0.099)
Age		-0.011*	-0.013**
		(0.006)	(0.006)
Female		-0.141*	-0.159**
		(0.081)	(0.072)
		Expenditure per month, excluding rent - baseline: £ 400 to 800	
< £ 400		-0.268***	-0.268***
		(0.099)	(0.082)
> £ 800		-0.255*	-0.287***
		(0.131)	(0.107)
		Ethnicity - baseline: Others	
Asian		-0.128	-0.089
		(0.100)	(0.083)
White		0.028	0.006
		(0.117)	(0.105)
		Session - baseline: Session 1	
Session 2		0.036	0.090
		(0.092)	(0.085)
Session 3		0.067	0.107
		(0.098)	(0.102)
Session 4		-0.091	-0.037
		(0.107)	(0.090)
Control	*No*	*Yes*	*Yes*
Obs.	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B20: Determinants of Conditional Cooperation in the Fehr-Schmidt model: Asymmetric games & CCEI > 0.80

	Ahead	Ahead	Ahead	Behind	Behind	Behind
DAMAGE	0.168*	0.143*	0.139	0.939***	0.851***	0.868***
	(0.093)	(0.087)	(0.086)	(0.325)	(0.300)	(0.294)
GAIN	-0.574***	-0.619***	-0.614***	-0.193**	-0.205**	-0.204**
	(0.192)	(0.194)	(0.191)	(0.089)	(0.086)	(0.086)
EFF	-0.444**	-0.449***	-0.436**	-0.048	-0.039	-0.040
	(0.175)	(0.174)	(0.170)	(0.052)	(0.050)	(0.050)
$\hat{\beta}$			0.202***			0.466***
			(0.070)			(0.085)
CCEI			0.020			0.044
			(0.375)			(0.364)
Round	-0.002	-0.003	-0.003	-0.002	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			the measures of Falk et al. (2018)			
P.Reciprocity		0.045**	0.038*		0.040	0.034
		(0.023)	(0.022)		(0.035)	(0.025)
N.Reciprocity		-0.038	-0.031		-0.059**	-0.053**
		(0.025)	(0.024)		(0.030)	(0.025)
Patience		0.008	0.005		-0.001	-0.012
		(0.025)	(0.024)		(0.031)	(0.025)
Risk taking		0.012	0.014		0.036	0.037
		(0.024)	(0.023)		(0.033)	(0.025)
Trust		0.016	0.008		0.032	0.003
		(0.020)	(0.019)		(0.021)	(0.014)
Math		0.019	0.023		0.018	0.034*
		(0.022)	(0.021)		(0.020)	(0.018)
Procrastination		-0.014	-0.009		-0.007	-0.011
		(0.029)	(0.028)		(0.035)	(0.026)
			Big-Five personality using Hahn et al. (2012)			
FiveC		-0.012	-0.006		0.014	0.019
		(0.023)	(0.021)		(0.027)	(0.020)
FiveA		-0.018	-0.005		-0.027	-0.008
		(0.026)	(0.024)		(0.026)	(0.020)
FiveE		-0.023	-0.021		-0.030	-0.027**
		(0.015)	(0.015)		(0.020)	(0.014)
FiveN		0.016	0.005		0.040*	0.019
		(0.016)	(0.016)		(0.021)	(0.018)
FiveO		0.039*	0.035		0.003	-0.010
		(0.023)	(0.023)		(0.022)	(0.020)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
			Demographic characteristics			
Politics		-0.020	-0.020		-0.022*	-0.017
		(0.013)	(0.013)		(0.013)	(0.012)
Economics		-0.117*	-0.120**		-0.088	-0.077
		(0.061)	(0.058)		(0.072)	(0.051)
Age		-0.000	-0.001		-0.002	-0.004
		(0.004)	(0.003)		(0.005)	(0.004)
Female		-0.034	-0.035		-0.101*	-0.110***
		(0.045)	(0.043)		(0.053)	(0.035)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		-0.158***	-0.161***		-0.028	-0.023
		(0.056)	(0.052)		(0.064)	(0.048)
$>$ £ 800		-0.112	-0.131**		0.017	-0.007
		(0.073)	(0.066)		(0.074)	(0.058)
			Ethnicity - baseline: Others			
Asian		-0.096*	-0.085*		-0.060	-0.035
		(0.050)	(0.048)		(0.062)	(0.053)
White		-0.063	-0.077		0.005	-0.015
		(0.055)	(0.052)		(0.076)	(0.060)
			Session - baseline: Session 1			
Session 2		0.072	0.107**		0.012	0.110**
		(0.047)	(0.047)		(0.058)	(0.054)
Session 3		0.099	0.118*		0.064	0.110*
		(0.061)	(0.064)		(0.057)	(0.059)
Session 4		-0.009	0.027		0.031	0.114**
		(0.054)	(0.049)		(0.075)	(0.058)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	688	664	664	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B21: Determinants of Free Riding in the Fehr-Schmidt model: Symmetric games & CCEI > 0.80

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
DAMAGE	-0.181**	-0.174**	-0.171**
	(0.087)	(0.088)	(0.087)
GAIN	0.184***	0.184***	0.183***
	(0.063)	(0.064)	(0.064)
EFF	-0.068	-0.053	-0.054
	(0.065)	(0.066)	(0.065)
$\hat{\beta}$			-0.518***
			(0.121)
CCEI			0.956
			(0.691)
Round	-0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		-0.124**	-0.121***
		(0.051)	(0.044)
N.Reciprocity		0.048	0.041
		(0.059)	(0.049)
Patience		-0.066	-0.066
		(0.050)	(0.049)
Risk taking		-0.048	-0.061
		(0.049)	(0.041)
Trust		-0.031	0.001
		(0.041)	(0.037)
Math		0.046	0.043
		(0.045)	(0.039)
Procrastination		0.055	0.048
		(0.054)	(0.048)
	Big-Five personality using Hahn et al. (2012)		
FiveC		0.046	0.018
		(0.046)	(0.041)
FiveA		0.016	0.001
		(0.051)	(0.045)
FiveE		0.098***	0.101***
		(0.035)	(0.029)
FiveN		-0.017	0.000
		(0.031)	(0.029)
FiveO		-0.037	-0.013
		(0.050)	(0.047)

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
		Demographic characteristics	
Politics		0.058**	0.062***
		(0.024)	(0.021)
Economics		0.102	0.059
		(0.133)	(0.111)
Age		0.014**	0.017***
		(0.006)	(0.006)
Female		0.155*	0.185**
		(0.084)	(0.076)
		Expenditure per month, excluding rent - baseline: £ 400 to 800	
< £ 400		0.262***	0.264***
		(0.098)	(0.080)
> £ 800		0.136	0.169
		(0.136)	(0.114)
		Ethnicity - baseline: Others	
Asian		0.138	0.090
		(0.103)	(0.087)
White		-0.033	-0.007
		(0.114)	(0.106)
		Session - baseline: Session 1	
Session 2		-0.026	-0.077
		(0.096)	(0.089)
Session 3		-0.140	-0.183*
		(0.107)	(0.110)
Session 4		0.063	0.003
		(0.112)	(0.095)
Control	*No*	*Yes*	*Yes*
Obs.	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B22: Determinants of Free Riding in the Fehr-Schmidt model: Asymmetric games & CCEI > 0.80

	Ahead	Ahead	Ahead	Behind	Behind	Behind
DAMAGE	-0.138	-0.111	-0.112	-0.610*	-0.600*	-0.601*
	(0.110)	(0.112)	(0.113)	(0.339)	(0.345)	(0.337)
GAIN	0.187	0.205	0.207	0.185*	0.183*	0.179*
	(0.170)	(0.177)	(0.178)	(0.098)	(0.099)	(0.098)
EFF	-0.442***	-0.414***	-0.415***	-0.103**	-0.096**	-0.095**
	(0.142)	(0.142)	(0.142)	(0.045)	(0.045)	(0.045)
$\hat{\beta}$			-0.557***			-0.540***
			(0.110)			(0.097)
CCEI			0.139			0.197
			(0.600)			(0.423)
Round	-0.001	-0.001	-0.001	0.003	0.003	0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			the measures of Falk et al. (2018)			
P.Reciprocity		-0.097**	-0.084**		-0.058	-0.041
		(0.047)	(0.039)		(0.044)	(0.033)
N.Reciprocity		0.031	0.015		0.075*	0.067**
		(0.052)	(0.045)		(0.041)	(0.034)
Patience		-0.043	-0.036		-0.002	0.004
		(0.047)	(0.043)		(0.036)	(0.031)
Risk taking		-0.075	-0.077**		-0.009	-0.007
		(0.047)	(0.038)		(0.041)	(0.032)
Trust		-0.009	0.013		-0.021	0.009
		(0.037)	(0.031)		(0.026)	(0.019)
Math		-0.010	-0.017		-0.004	-0.022
		(0.042)	(0.035)		(0.027)	(0.023)
Procrastination		0.066	0.051		-0.007	-0.006
		(0.053)	(0.047)		(0.047)	(0.039)
			Big-Five personality using Hahn et al. (2012)			
FiveC		0.082*	0.056		0.001	-0.006
		(0.044)	(0.041)		(0.036)	(0.028)
FiveA		-0.027	-0.054		0.008	-0.014
		(0.054)	(0.044)		(0.034)	(0.029)
FiveE		0.061*	0.058*		0.035	0.028
		(0.034)	(0.030)		(0.027)	(0.020)
FiveN		-0.023	0.004		-0.038	-0.015
		(0.030)	(0.027)		(0.027)	(0.023)
FiveO		-0.053	-0.039		-0.021	-0.004
		(0.048)	(0.044)		(0.027)	(0.024)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
			Demographic characteristics			
Politics		0.068***	0.067***		0.018	0.012
		(0.023)	(0.021)		(0.016)	(0.015)
Economics		0.024	0.009		0.039	0.009
		(0.117)	(0.100)		(0.098)	(0.079)
Age		0.013*	0.015**		0.007	0.010*
		(0.006)	(0.006)		(0.006)	(0.005)
Female		0.117	0.129*		0.139**	0.156***
		(0.085)	(0.076)		(0.064)	(0.047)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		0.214**	0.216***		-0.030	-0.044
		(0.098)	(0.080)		(0.072)	(0.056)
$>$ £ 800		0.200	0.229*		-0.071	-0.068
		(0.142)	(0.121)		(0.090)	(0.069)
			Ethnicity - baseline: Others			
Asian		0.193*	0.151*		0.086	0.053
		(0.107)	(0.091)		(0.082)	(0.077)
White		0.089	0.125		0.023	0.053
		(0.114)	(0.095)		(0.096)	(0.081)
			Session - baseline: Session 1			
Session 2		-0.104	-0.191**		-0.058	-0.169**
		(0.102)	(0.091)		(0.078)	(0.069)
Session 3		-0.170*	-0.216**		-0.065	-0.109
		(0.098)	(0.097)		(0.081)	(0.084)
Session 4		0.016	-0.057		-0.063	-0.150*
		(0.104)	(0.078)		(0.095)	(0.078)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	688	664	664	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B23: Determinants of First-mover's Cooperation in the Fehr-Schmidt model: Symmetric games & CCEI > 0.80

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
RISK	-0.053	-0.065	-0.066
	(0.061)	(0.065)	(0.065)
TEMPT	-0.225**	-0.229**	-0.228**
	(0.100)	(0.097)	(0.096)
EFF	0.143**	0.128**	0.129**
	(0.057)	(0.053)	(0.052)
$\hat{\beta}$			0.310***
			(0.093)
CCEI			-0.734
			(0.454)
Round	-0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		-0.007	-0.012
		(0.034)	(0.033)
N.Reciprocity		-0.043	-0.037
		(0.041)	(0.038)
Patience		-0.002	0.008
		(0.031)	(0.033)
Risk taking		0.064*	0.061**
		(0.036)	(0.027)
Trust		0.030	0.013
		(0.031)	(0.030)
Math		-0.095**	-0.090**
		(0.039)	(0.036)
Procrastination		-0.014	-0.024
		(0.042)	(0.035)
	Big-Five personality using Hahn et al. (2012)		
FiveC		-0.039	-0.039
		(0.033)	(0.026)
FiveA		-0.002	0.011
		(0.040)	(0.037)
FiveE		-0.086***	-0.090***
		(0.028)	(0.023)
FiveN		0.003	-0.010
		(0.024)	(0.022)
FiveO		0.042	0.019
		(0.033)	(0.031)

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
		Demographic characteristics	
Politics		-0.043**	-0.041**
		(0.018)	(0.017)
Economics		0.016	0.051
		(0.090)	(0.087)
Age		-0.000	-0.002
		(0.005)	(0.005)
Female		-0.080	-0.125*
		(0.066)	(0.068)
		Expenditure per month, excluding rent - baseline: £ 400 to 800	
< £ 400		0.002	-0.004
		(0.076)	(0.073)
> £ 800		0.031	0.012
		(0.076)	(0.067)
		Ethnicity - baseline: Others	
Asian		0.174***	0.199***
		(0.061)	(0.054)
White		0.137**	0.141**
		(0.069)	(0.067)
		Session - baseline: Session 1	
Session 2		-0.013	0.027
		(0.072)	(0.068)
Session 3		0.144*	0.159**
		(0.079)	(0.080)
Session 4		-0.124	-0.094
		(0.094)	(0.079)
Control	*No*	*Yes*	*Yes*
Obs.	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B24: Determinants of First-mover's Cooperation in the Fehr-Schmidt model: Asymmetric games & CCEI > 0.80

	Ahead	Ahead	Ahead	Behind	Behind	Behind
RISK	0.395	0.343	0.340	0.061	0.075*	0.074*
	(0.274)	(0.291)	(0.289)	(0.040)	(0.040)	(0.039)
TEMPT	-0.274	-0.332	-0.341	-0.224**	-0.227**	-0.226**
	(0.276)	(0.296)	(0.300)	(0.101)	(0.093)	(0.093)
EFF	0.526**	0.484*	0.480*	0.037	0.045	0.044
	(0.245)	(0.251)	(0.248)	(0.041)	(0.040)	(0.039)
$\hat{\beta}$			0.529***			0.085
			(0.099)			(0.061)
CCEI			0.285			0.211
			(0.417)			(0.337)
Round	-0.001	-0.001	-0.001	-0.002	-0.002*	-0.002*
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
			the measures of Falk et al. (2018)			
P.Reciprocity		0.044	0.024		0.027	0.019
		(0.035)	(0.028)		(0.018)	(0.017)
N.Reciprocity		-0.007	0.004		-0.020	-0.012
		(0.040)	(0.033)		(0.018)	(0.019)
Patience		0.005	-0.001		0.012	0.011
		(0.034)	(0.031)		(0.016)	(0.016)
Risk taking		0.035	0.022		-0.008	-0.007
		(0.037)	(0.030)		(0.013)	(0.013)
Trust		0.009	-0.019		0.009	0.008
		(0.024)	(0.018)		(0.012)	(0.012)
Math		-0.011	-0.008		-0.031*	-0.029*
		(0.028)	(0.024)		(0.018)	(0.016)
Procrastination		-0.015	-0.030		-0.040	-0.037
		(0.048)	(0.033)		(0.025)	(0.022)
			Big-Five personality using Hahn et al. (2012)			
FiveC		-0.054	-0.041		-0.030*	-0.023
		(0.036)	(0.028)		(0.017)	(0.015)
FiveA		0.024	0.040		-0.016	-0.010
		(0.031)	(0.029)		(0.018)	(0.017)
FiveE		-0.073***	-0.063***		-0.014	-0.015*
		(0.025)	(0.020)		(0.009)	(0.009)
FiveN		0.031	0.022		0.014	0.009
		(0.022)	(0.019)		(0.010)	(0.010)
FiveO		0.033	0.028		0.035**	0.033**
		(0.029)	(0.025)		(0.015)	(0.014)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
			Demographic characteristics			
Politics		-0.010	0.000		-0.011	-0.009
		(0.017)	(0.014)		(0.010)	(0.010)
Economics		-0.049	-0.097		-0.033	-0.041
		(0.109)	(0.106)		(0.044)	(0.045)
Age		-0.010	-0.012***		-0.005*	-0.006**
		(0.007)	(0.005)		(0.003)	(0.003)
Female		0.076	0.031		-0.081***	-0.079***
		(0.064)	(0.044)		(0.029)	(0.029)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		0.006	0.028		0.006	0.004
		(0.062)	(0.049)		(0.033)	(0.034)
$>$ £ 800		0.164	0.129		-0.019	-0.020
		(0.138)	(0.093)		(0.036)	(0.034)
			Ethnicity - baseline: Others			
Asian		-0.034	-0.031		-0.008	-0.007
		(0.081)	(0.061)		(0.033)	(0.035)
White		-0.043	-0.102		0.028	0.010
		(0.085)	(0.070)		(0.040)	(0.038)
			Session - baseline: Session 1			
Session 2		0.086	0.161***		0.008	0.020
		(0.067)	(0.061)		(0.033)	(0.029)
Session 3		0.191**	0.213***		0.038	0.042
		(0.081)	(0.076)		(0.039)	(0.041)
Session 4		0.131	0.193***		-0.072	-0.047
		(0.087)	(0.061)		(0.044)	(0.033)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	688	664	664	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B25: Determinants of Conditional Cooperation in the Charness-Rabin model: Symmetric games

	All	All	All
DAMAGE	0.216***	0.208***	0.206***
	(0.075)	(0.076)	(0.075)
GAIN	-0.140**	-0.137**	-0.136**
	(0.058)	(0.059)	(0.058)
EFF	-0.008	-0.021	-0.022
	(0.066)	(0.066)	(0.065)
$\hat{\lambda}$			-0.238*
			(0.136)
$\hat{\delta}$			0.349***
			(0.116)
CCEI			-0.039
			(0.172)
Round	0.001	0.001	0.001
	(0.002)	(0.001)	(0.001)
	the measures of Falk et al. (2018)		
P.Reciprocity		0.136***	0.102**
		(0.045)	(0.041)
N.Reciprocity		-0.014	-0.023
		(0.049)	(0.044)
Patience		0.027	0.058
		(0.047)	(0.045)
Risk taking		0.065	0.074**
		(0.043)	(0.038)
Trust		0.030	-0.006
		(0.032)	(0.033)
Math		-0.034	-0.034
		(0.036)	(0.032)
Procrastination		-0.032	-0.002
		(0.047)	(0.044)
	Big-Five personality using Hahn et al. (2012)		
FiveC		-0.021	0.012
		(0.040)	(0.038)
FiveA		-0.012	-0.001
		(0.047)	(0.041)
FiveE		-0.040	-0.079***
		(0.028)	(0.026)
FiveN		0.008	-0.024
		(0.032)	(0.031)

	All	All	All
FiveO		0.010	-0.025
		(0.043)	(0.040)
	Demographic characteristics		
Politics		-0.054***	-0.070***
		(0.019)	(0.016)
Economics		-0.275**	-0.199*
		(0.109)	(0.113)
Age		-0.006	-0.009
		(0.006)	(0.006)
Female		-0.047	-0.105*
		(0.067)	(0.058)
	Expenditure per month, excluding rent - baseline: £ 400 to 800		
$<$ £ 400		-0.174*	-0.180**
		(0.090)	(0.078)
$>$ £ 800		-0.283***	-0.242**
		(0.105)	(0.102)
	Ethnicity - baseline: Others		
Asian		-0.171*	-0.087
		(0.088)	(0.074)
White		-0.099	-0.056
		(0.096)	(0.097)
	Session - baseline: Session 1		
Session 2		0.040	0.128
		(0.085)	(0.078)
Session 3		0.141	0.217**
		(0.089)	(0.086)
Session 4		-0.087	0.015
		(0.096)	(0.089)
Control	*No*	*Yes*	*Yes*
Obs.	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B26: Determinants of Conditional Cooperation in the Charness-Rabin model: Asymmetric games

	Ahead	Ahead	Ahead	Behind	Behind	Behind
DAMAGE	0.099	0.074	0.069	0.895***	0.829***	0.851***
	(0.089)	(0.085)	(0.084)	(0.313)	(0.297)	(0.297)
GAIN	-0.563***	-0.594***	-0.587***	-0.151*	-0.158**	-0.152*
	(0.167)	(0.168)	(0.165)	(0.082)	(0.080)	(0.079)
EFF	-0.390***	-0.397***	-0.385***	-0.025	-0.019	-0.016
	(0.149)	(0.150)	(0.147)	(0.053)	(0.053)	(0.052)
$\hat{\lambda}$			-0.176***			-0.326***
			(0.064)			(0.073)
$\hat{\delta}$			0.086			0.040
			(0.061)			(0.066)
CCEI			0.154			-0.009
			(0.128)			(0.104)
Round	-0.002	-0.003	-0.003	-0.003	-0.003*	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			the measures of Falk et al. (2018)			
P.Reciprocity		0.059***	0.041**		0.038	0.024
		(0.022)	(0.021)		(0.028)	(0.024)
N.Reciprocity		-0.012	-0.011		-0.020	-0.024
		(0.023)	(0.020)		(0.024)	(0.020)
Patience		-0.003	0.002		-0.003	0.001
		(0.024)	(0.022)		(0.026)	(0.021)
Risk taking		0.029	0.029		0.062**	0.056**
		(0.022)	(0.020)		(0.030)	(0.024)
Trust		0.007	0.001		0.024	-0.012
		(0.017)	(0.019)		(0.018)	(0.017)
Math		-0.003	0.002		0.020	0.018
		(0.018)	(0.017)		(0.016)	(0.013)
Procrastination		-0.013	-0.001		-0.018	-0.012
		(0.023)	(0.023)		(0.029)	(0.022)
			Big-Five personality using Hahn et al. (2012)			
FiveC		0.002	0.013		-0.009	0.004
		(0.022)	(0.020)		(0.024)	(0.022)
FiveA		-0.008	0.004		0.007	0.005
		(0.024)	(0.020)		(0.024)	(0.020)
FiveE		-0.007	-0.013		-0.034**	-0.045***
		(0.013)	(0.014)		(0.017)	(0.015)
FiveN		0.004	-0.012		0.035**	0.016
		(0.013)	(0.013)		(0.017)	(0.014)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
FiveO		0.014	0.003		0.005	-0.008
		(0.021)	(0.019)		(0.017)	(0.014)
			Demographic characteristics			
Politics		-0.026***	-0.029***		-0.019*	-0.025***
		(0.010)	(0.010)		(0.010)	(0.009)
Economics		-0.140**	-0.124**		-0.109	-0.041
		(0.064)	(0.056)		(0.069)	(0.063)
Age		0.001	0.000		0.000	-0.002
		(0.003)	(0.003)		(0.004)	(0.003)
Female		-0.011	-0.014		-0.079*	-0.092**
		(0.037)	(0.037)		(0.042)	(0.037)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		-0.123***	-0.116***		0.001	0.018
		(0.046)	(0.045)		(0.051)	(0.040)
$>$ £ 800		-0.147***	-0.136***		-0.015	0.003
		(0.055)	(0.052)		(0.062)	(0.047)
			Ethnicity - baseline: Others			
Asian		-0.084*	-0.057		-0.103**	-0.045
		(0.045)	(0.041)		(0.047)	(0.042)
White		-0.073	-0.075*		-0.059	-0.048
		(0.047)	(0.043)		(0.060)	(0.053)
			Session - baseline: Session 1			
Session 2		0.060	0.114***		0.015	0.067
		(0.041)	(0.041)		(0.051)	(0.049)
Session 3		0.072	0.108*		0.064	0.080
		(0.052)	(0.055)		(0.050)	(0.055)
Session 4		-0.049	0.015		0.042	0.094*
		(0.053)	(0.050)		(0.064)	(0.056)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	856	832	832	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B27: Determinants of Free Riding in the Charness-Rabin model: Symmetric games

	All	All	All
DAMAGE	-0.166*	-0.160*	-0.159*
	(0.087)	(0.088)	(0.088)
GAIN	0.167***	0.166***	0.167***
	(0.057)	(0.057)	(0.058)
EFF	-0.047	-0.034	-0.034
	(0.057)	(0.057)	(0.058)
$\hat{\lambda}$			0.329**
			(0.137)
$\hat{\delta}$			-0.299**
			(0.126)
CCEI			0.215
			(0.200)
Round	-0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		-0.147***	-0.109**
		(0.048)	(0.045)
N.Reciprocity		0.019	0.033
		(0.055)	(0.048)
Patience		-0.051	-0.079*
		(0.050)	(0.048)
Risk taking		-0.055	-0.070*
		(0.044)	(0.037)
Trust		-0.028	0.028
		(0.033)	(0.033)
Math		0.022	0.029
		(0.035)	(0.028)
Procrastination		0.058	0.028
		(0.049)	(0.043)
	Big-Five personality using Hahn et al. (2012)		
FiveC		0.070	0.027
		(0.047)	(0.043)
FiveA		-0.024	-0.034
		(0.051)	(0.044)
FiveE		0.041	0.082***
		(0.031)	(0.027)
FiveN		-0.006	0.021
		(0.031)	(0.028)

	All	All	All
FiveO		-0.034	-0.008
		(0.044)	(0.040)
	Demographic characteristics		
Politics		0.031	0.046***
		(0.021)	(0.017)
Economics		0.262**	0.148
		(0.117)	(0.121)
Age		0.010	0.014**
		(0.007)	(0.006)
Female		0.047	0.131**
		(0.073)	(0.062)
	Expenditure per month, excluding rent - baseline: £ 400 to 800		
$<$ £ 400		0.166*	0.179**
		(0.091)	(0.075)
$>$ £ 800		0.117	0.075
		(0.132)	(0.115)
	Ethnicity - baseline: Others		
Asian		0.170**	0.058
		(0.086)	(0.072)
White		0.084	0.041
		(0.091)	(0.093)
	Session - baseline: Session 1		
Session 2		-0.041	-0.120
		(0.091)	(0.082)
Session 3		-0.143	-0.209**
		(0.098)	(0.093)
Session 4		0.072	-0.032
		(0.094)	(0.084)
Control	*No*	*Yes*	*Yes*
Obs.	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B28: Determinants of Free Riding in the Charness-Rabin model: Asymmetric games

	Ahead	Ahead	Ahead	Behind	Behind	Behind
DAMAGE	-0.134	-0.111	-0.111	-0.545*	-0.531*	-0.535*
	(0.096)	(0.097)	(0.098)	(0.316)	(0.319)	(0.315)
GAIN	0.262*	0.278*	0.283*	0.162*	0.160*	0.156*
	(0.150)	(0.154)	(0.156)	(0.086)	(0.088)	(0.086)
EFF	-0.293**	-0.265**	-0.267**	-0.075	-0.067	-0.068
	(0.131)	(0.132)	(0.134)	(0.047)	(0.048)	(0.048)
$\hat{\lambda}$			0.447***			0.442***
			(0.127)			(0.098)
$\hat{\delta}$			-0.135			-0.061
			(0.124)			(0.095)
CCEI			-0.148			0.169
			(0.196)			(0.152)
Round	-0.001	-0.002	-0.001	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			the measures of Falk et al. (2018)			
P.Reciprocity		-0.127***	-0.086**		-0.057	-0.020
		(0.046)	(0.041)		(0.040)	(0.034)
N.Reciprocity		0.022	0.019		0.022	0.030
		(0.052)	(0.044)		(0.035)	(0.030)
Patience		-0.037	-0.051		-0.013	-0.022
		(0.047)	(0.043)		(0.036)	(0.032)
Risk taking		-0.068	-0.069**		-0.031	-0.017
		(0.043)	(0.035)		(0.042)	(0.034)
Trust		0.008	0.032		-0.025	0.027
		(0.032)	(0.032)		(0.026)	(0.022)
Math		0.012	0.010		-0.034	-0.034
		(0.034)	(0.028)		(0.027)	(0.021)
Procrastination		0.046	0.015		0.029	0.016
		(0.046)	(0.042)		(0.042)	(0.033)
			Big-Five personality using Hahn et al. (2012)			
FiveC		0.069	0.034		0.066*	0.039
		(0.045)	(0.041)		(0.036)	(0.031)
FiveA		-0.044	-0.063		-0.036	-0.035
		(0.051)	(0.041)		(0.032)	(0.027)
FiveE		0.010	0.030		0.032	0.049**
		(0.032)	(0.030)		(0.024)	(0.020)
FiveN		-0.011	0.024		-0.026	-0.008
		(0.029)	(0.026)		(0.025)	(0.021)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
FiveO		-0.026	-0.003		-0.030	-0.015
		(0.039)	(0.033)		(0.025)	(0.021)
			Demographic characteristics			
Politics		0.060***	0.067***		0.000	0.005
		(0.020)	(0.017)		(0.016)	(0.014)
Economics		0.081	-0.005		0.138	0.017
		(0.113)	(0.107)		(0.107)	(0.100)
Age		0.011	0.013**		0.004	0.008*
		(0.006)	(0.006)		(0.005)	(0.004)
Female		0.051	0.077		0.090	0.139***
		(0.073)	(0.068)		(0.063)	(0.053)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		0.106	0.102		-0.041	-0.062
		(0.086)	(0.071)		(0.065)	(0.051)
$>$ £ 800		0.166	0.137		-0.107	-0.149*
		(0.129)	(0.122)		(0.115)	(0.087)
			Ethnicity - baseline: Others			
Asian		0.149*	0.052		0.164**	0.070
		(0.089)	(0.079)		(0.071)	(0.066)
White		0.123	0.107		0.091	0.072
		(0.093)	(0.083)		(0.083)	(0.072)
			Session - baseline: Session 1			
Session 2		-0.058	-0.163*		-0.065	-0.127**
		(0.089)	(0.084)		(0.070)	(0.063)
Session 3		-0.077	-0.138		-0.045	-0.056
		(0.095)	(0.091)		(0.077)	(0.081)
Session 4		0.104	-0.015		-0.083	-0.138*
		(0.098)	(0.084)		(0.087)	(0.078)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	856	832	832	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B29: Determinants of First-mover's Cooperation in the Charness-Rabin model: Symmetric games

	All	All	All
RISK	-0.010	-0.017	-0.020
	(0.055)	(0.060)	(0.061)
TEMPT	-0.209**	-0.215**	-0.219**
	(0.089)	(0.091)	(0.092)
EFF	0.109**	0.100*	0.101*
	(0.053)	(0.054)	(0.054)
$\hat{\lambda}$			-0.262**
			(0.114)
$\hat{\delta}$			0.132
			(0.110)
CCEI			-0.151
			(0.164)
Round	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		0.027	-0.014
		(0.036)	(0.038)
N.Reciprocity		-0.021	-0.017
		(0.043)	(0.039)
Patience		-0.007	0.015
		(0.032)	(0.032)
Risk taking		0.073**	0.065**
		(0.034)	(0.030)
Trust		0.029	0.003
		(0.028)	(0.030)
Math		-0.074**	-0.078**
		(0.035)	(0.032)
Procrastination		-0.007	0.007
		(0.046)	(0.044)
	Big-Five personality using Hahn et al. (2012)		
FiveC		-0.043	-0.024
		(0.037)	(0.034)
FiveA		0.023	0.041
		(0.042)	(0.040)
FiveE		-0.050*	-0.074***
		(0.028)	(0.027)
FiveN		-0.012	-0.038
		(0.027)	(0.026)

	All	All	All
FiveO		0.018	-0.003
		(0.033)	(0.033)
	Demographic characteristics		
Politics		-0.021	-0.027*
		(0.017)	(0.015)
Economics		-0.055	0.015
		(0.089)	(0.104)
Age		-0.007	-0.010
		(0.006)	(0.007)
Female		0.021	-0.040
		(0.066)	(0.065)
	Expenditure per month, excluding rent - baseline: £ 400 to 800		
$<$ £ 400		0.058	0.051
		(0.065)	(0.060)
$>$ £ 800		0.030	0.065
		(0.074)	(0.078)
	Ethnicity - baseline: Others		
Asian		0.026	0.096
		(0.077)	(0.066)
White		0.001	0.034
		(0.078)	(0.074)
	Session - baseline: Session 1		
Session 2		0.023	0.075
		(0.070)	(0.069)
Session 3		0.029	0.061
		(0.084)	(0.096)
Session 4		-0.060	-0.007
		(0.085)	(0.081)
Control	*No*	*Yes*	*Yes*
Obs.	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B30: Determinants of First-mover's Cooperation in the Charness-Rabin model: Asymmetric games

	Ahead	Ahead	Ahead	Behind	Behind	Behind
RISK	0.307	0.245	0.248	0.053	0.065*	0.065*
	(0.277)	(0.286)	(0.286)	(0.036)	(0.037)	(0.037)
TEMPT	-0.362	-0.399	-0.404	-0.192**	-0.196**	-0.200**
	(0.269)	(0.275)	(0.278)	(0.085)	(0.083)	(0.084)
EFF	0.406*	0.343	0.343	0.025	0.030	0.031
	(0.234)	(0.238)	(0.237)	(0.036)	(0.036)	(0.036)
$\hat{\lambda}$			-0.449***			-0.143*
			(0.100)			(0.078)
$\hat{\delta}$			0.145*			-0.016
			(0.085)			(0.074)
CCEI			0.077			-0.064
			(0.154)			(0.128)
Round	0.000	0.000	0.000	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
			the measures of Falk et al. (2018)			
P.Reciprocity		0.045	0.006		0.033	0.031
		(0.036)	(0.032)		(0.022)	(0.021)
N.Reciprocity		0.036	0.031		0.006	0.006
		(0.039)	(0.030)		(0.024)	(0.022)
Patience		0.004	0.038		0.013	0.024
		(0.035)	(0.035)		(0.019)	(0.020)
Risk taking		0.037	0.039		0.018	0.018
		(0.035)	(0.027)		(0.018)	(0.020)
Trust		0.014	-0.038		0.009	-0.006
		(0.021)	(0.025)		(0.014)	(0.016)
Math		-0.014	-0.017		-0.043**	-0.046***
		(0.025)	(0.020)		(0.018)	(0.017)
Procrastination		-0.044	-0.040		-0.036	-0.032
		(0.041)	(0.032)		(0.025)	(0.024)
			Big-Five personality using Hahn et al. (2012)			
FiveC		-0.056	-0.042		-0.017	-0.007
		(0.036)	(0.028)		(0.019)	(0.019)
FiveA		0.033	0.040		0.002	-0.001
		(0.033)	(0.025)		(0.023)	(0.022)
FiveE		-0.049**	-0.072***		-0.016	-0.025*
		(0.024)	(0.022)		(0.012)	(0.013)
FiveN		0.006	-0.013		0.023*	0.016
		(0.023)	(0.021)		(0.012)	(0.012)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
FiveO		0.014	-0.005		0.031*	0.030
		(0.027)	(0.025)		(0.018)	(0.019)
			Demographic characteristics			
Politics		-0.011	-0.019		-0.004	-0.005
		(0.014)	(0.013)		(0.010)	(0.009)
Economics		-0.184*	-0.138		-0.093	-0.079
		(0.104)	(0.104)		(0.072)	(0.078)
Age		-0.013**	-0.018***		-0.005	-0.007**
		(0.006)	(0.005)		(0.003)	(0.003)
Female		0.063	0.020		-0.049	-0.061*
		(0.055)	(0.049)		(0.032)	(0.032)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		0.034	0.077*		0.081***	0.085***
		(0.059)	(0.046)		(0.029)	(0.029)
$>$ £ 800		0.097	0.192*		-0.002	0.003
		(0.111)	(0.105)		(0.026)	(0.029)
			Ethnicity - baseline: Others			
Asian		-0.098	-0.032		-0.022	0.003
		(0.079)	(0.065)		(0.039)	(0.036)
White		-0.117	-0.118*		0.006	0.010
		(0.079)	(0.064)		(0.045)	(0.041)
			Session - baseline: Session 1			
Session 2		0.046	0.120**		-0.019	-0.014
		(0.064)	(0.060)		(0.041)	(0.039)
Session 3		0.078	0.102		-0.012	-0.019
		(0.071)	(0.072)		(0.043)	(0.050)
Session 4		0.072	0.168***		-0.076	-0.076*
		(0.075)	(0.057)		(0.049)	(0.042)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	856	832	832	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B31: Determinants of Conditional Cooperation in the Charness-Rabin model: Symmetric games & CCEI > 0.80

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
DAMAGE	0.168**	0.161**	0.159**
	(0.077)	(0.079)	(0.078)
GAIN	-0.188***	-0.191***	-0.189***
	(0.064)	(0.065)	(0.065)
EFF	0.004	-0.013	-0.013
	(0.073)	(0.076)	(0.074)
$\hat{\lambda}$			-0.099
			(0.144)
$\hat{\delta}$			0.406***
			(0.110)
CCEI			-0.131
			(0.616)
Round	0.001	0.000	0.000
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		0.115**	0.102**
		(0.047)	(0.043)
N.Reciprocity		-0.029	-0.048
		(0.053)	(0.043)
Patience		0.051	0.061
		(0.049)	(0.043)
Risk taking		0.048	0.064*
		(0.048)	(0.037)
Trust		0.041	0.018
		(0.040)	(0.037)
Math		-0.039	-0.034
		(0.046)	(0.040)
Procrastination		-0.056	-0.034
		(0.053)	(0.046)
	Big-Five personality using Hahn et al. (2012)		
FiveC		-0.024	0.006
		(0.041)	(0.039)
FiveA		-0.027	-0.022
		(0.048)	(0.041)
FiveE		-0.071**	-0.092***
		(0.031)	(0.026)
FiveN		0.003	-0.025
		(0.032)	(0.028)

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
FiveO		0.009	-0.042
		(0.049)	(0.046)
	Demographic characteristics		
Politics		-0.068***	-0.077***
		(0.024)	(0.020)
Economics		-0.180	-0.145
		(0.115)	(0.095)
Age		-0.011*	-0.011**
		(0.006)	(0.005)
Female		-0.141*	-0.180**
		(0.081)	(0.072)
	Expenditure per month, excluding rent - baseline: £ 400 to 800		
< £ 400		-0.268***	-0.299***
		(0.099)	(0.074)
> £ 800		-0.255*	-0.244**
		(0.131)	(0.117)
	Ethnicity - baseline: Others		
Asian		-0.128	-0.071
		(0.100)	(0.081)
White		0.028	0.094
		(0.117)	(0.099)
	Session - baseline: Session 1		
Session 2		0.036	0.137*
		(0.092)	(0.081)
Session 3		0.067	0.169*
		(0.098)	(0.089)
Session 4		-0.091	0.040
		(0.107)	(0.091)
Control	*No*	*Yes*	*Yes*
Obs.	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B32: Determinants of Conditional Cooperation in the Charness-Rabin model: Asymmetric games & CCEI > 0.80

	Ahead	Ahead	Ahead	Behind	Behind	Behind
DAMAGE	0.168*	0.143*	0.139	0.939***	0.851***	0.870***
	(0.093)	(0.087)	(0.086)	(0.325)	(0.300)	(0.296)
GAIN	-0.574***	-0.619***	-0.613***	-0.193**	-0.205**	-0.204**
	(0.192)	(0.194)	(0.191)	(0.089)	(0.086)	(0.085)
EFF	-0.444**	-0.449***	-0.432**	-0.048	-0.039	-0.040
	(0.175)	(0.174)	(0.170)	(0.052)	(0.050)	(0.050)
$\hat{\lambda}$			-0.129*			-0.362***
			(0.070)			(0.088)
$\hat{\delta}$			0.126*			0.095
			(0.068)			(0.086)
CCEI			0.258			0.060
			(0.402)			(0.377)
Round	-0.002	-0.003	-0.003	-0.002	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			the measures of Falk et al. (2018)			
P.Reciprocity		0.045**	0.038		0.040	0.052*
		(0.023)	(0.023)		(0.035)	(0.029)
N.Reciprocity		-0.038	-0.042*		-0.059**	-0.069***
		(0.025)	(0.022)		(0.030)	(0.024)
Patience		0.008	0.003		-0.001	-0.024
		(0.025)	(0.024)		(0.031)	(0.026)
Risk taking		0.012	0.016		0.036	0.043*
		(0.024)	(0.022)		(0.033)	(0.025)
Trust		0.016	0.006		0.032	-0.003
		(0.020)	(0.018)		(0.021)	(0.014)
Math		0.019	0.023		0.018	0.027
		(0.022)	(0.020)		(0.020)	(0.018)
Procrastination		-0.014	-0.003		-0.007	-0.005
		(0.029)	(0.027)		(0.035)	(0.025)
			Big-Five personality using Hahn et al. (2012)			
FiveC		-0.012	0.001		0.014	0.028
		(0.023)	(0.021)		(0.027)	(0.020)
FiveA		-0.018	-0.005		-0.027	-0.013
		(0.026)	(0.022)		(0.026)	(0.020)
FiveE		-0.023	-0.024		-0.030	-0.032**
		(0.015)	(0.015)		(0.020)	(0.014)
FiveN		0.016	-0.000		0.040*	0.016
		(0.016)	(0.015)		(0.021)	(0.017)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
FiveO		0.039*	0.026		0.003	-0.016
		(0.023)	(0.022)		(0.022)	(0.020)
			Demographic characteristics			
Politics		-0.020	-0.018		-0.022*	-0.013
		(0.013)	(0.013)		(0.013)	(0.012)
Economics		-0.117*	-0.120**		-0.088	-0.061
		(0.061)	(0.056)		(0.072)	(0.053)
Age		-0.000	-0.001		-0.002	-0.004
		(0.004)	(0.003)		(0.005)	(0.003)
Female		-0.034	-0.040		-0.101*	-0.098**
		(0.045)	(0.044)		(0.053)	(0.041)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		-0.158***	-0.172***		-0.028	-0.027
		(0.056)	(0.051)		(0.064)	(0.050)
$>$ £ 800		-0.112	-0.125*		0.017	-0.020
		(0.073)	(0.067)		(0.074)	(0.056)
			Ethnicity - baseline: Others			
Asian		-0.096*	-0.074		-0.060	-0.034
		(0.050)	(0.046)		(0.062)	(0.053)
White		-0.063	-0.055		0.005	0.005
		(0.055)	(0.049)		(0.076)	(0.060)
			Session - baseline: Session 1			
Session 2		0.072	0.134***		0.012	0.118**
		(0.047)	(0.046)		(0.058)	(0.054)
Session 3		0.099	0.139**		0.064	0.104*
		(0.061)	(0.063)		(0.057)	(0.059)
Session 4		-0.009	0.056		0.031	0.122**
		(0.054)	(0.050)		(0.075)	(0.060)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	688	664	664	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B33: Determinants of Free Riding in the Charness-Rabin model: Symmetric games & CCEI > 0.80

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
DAMAGE	-0.181**	-0.174**	-0.170**
	(0.087)	(0.088)	(0.087)
GAIN	0.184***	0.184***	0.183***
	(0.063)	(0.064)	(0.064)
EFF	-0.068	-0.053	-0.053
	(0.065)	(0.066)	(0.065)
$\hat{\lambda}$			0.126
			(0.141)
$\hat{\delta}$			-0.399***
			(0.109)
CCEI			0.255
			(0.660)
Round	-0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		-0.124**	-0.113**
		(0.051)	(0.045)
N.Reciprocity		0.048	0.072
		(0.059)	(0.047)
Patience		-0.066	-0.070
		(0.050)	(0.045)
Risk taking		-0.048	-0.069*
		(0.049)	(0.038)
Trust		-0.031	-0.003
		(0.041)	(0.037)
Math		0.046	0.047
		(0.045)	(0.038)
Procrastination		0.055	0.040
		(0.054)	(0.047)
	Big-Five personality using Hahn et al. (2012)		
FiveC		0.046	0.009
		(0.046)	(0.043)
FiveA		0.016	0.015
		(0.051)	(0.044)
FiveE		0.098***	0.113***
		(0.035)	(0.029)
FiveN		-0.017	0.004
		(0.031)	(0.027)

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
FiveO		-0.037	0.014
		(0.050)	(0.047)
	Demographic characteristics		
Politics		0.058**	0.066***
		(0.024)	(0.020)
Economics		0.102	0.051
		(0.133)	(0.107)
Age		0.014**	0.015***
		(0.006)	(0.006)
Female		0.155*	0.204***
		(0.084)	(0.073)
	Expenditure per month, excluding rent - baseline: £ 400 to 800		
< £ 400		0.262***	0.293***
		(0.098)	(0.071)
> £ 800		0.136	0.120
		(0.136)	(0.117)
	Ethnicity - baseline: Others		
Asian		0.138	0.070
		(0.103)	(0.084)
White		-0.033	-0.091
		(0.114)	(0.099)
	Session - baseline: Session 1		
Session 2		-0.026	-0.124
		(0.096)	(0.084)
Session 3		-0.140	-0.242**
		(0.107)	(0.094)
Session 4		0.063	-0.073
		(0.112)	(0.091)
Control	*No*	*Yes*	*Yes*
Obs.	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B34: Determinants of Free Riding in the Charness-Rabin model: Asymmetric games & CCEI > 0.80

	Ahead	Ahead	Ahead	Behind	Behind	Behind
DAMAGE	-0.138	-0.111	-0.114	-0.610*	-0.600*	-0.604*
	(0.110)	(0.112)	(0.114)	(0.339)	(0.345)	(0.342)
GAIN	0.187	0.205	0.209	0.185*	0.183*	0.182*
	(0.170)	(0.177)	(0.180)	(0.098)	(0.099)	(0.099)
EFF	-0.442***	-0.414***	-0.422***	-0.103**	-0.096**	-0.096**
	(0.142)	(0.142)	(0.144)	(0.045)	(0.045)	(0.045)
$\hat{\lambda}$			0.316**			0.453***
			(0.144)			(0.119)
$\hat{\delta}$			-0.253**			-0.069
			(0.117)			(0.111)
CCEI			-0.252			0.202
			(0.600)			(0.422)
Round	-0.001	-0.001	-0.001	0.003	0.003	0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			the measures of Falk et al. (2018)			
P.Reciprocity		-0.097**	-0.087**		-0.058	-0.060
		(0.047)	(0.044)		(0.044)	(0.037)
N.Reciprocity		0.031	0.046		0.075*	0.087***
		(0.052)	(0.044)		(0.041)	(0.032)
Patience		-0.043	-0.037		-0.002	0.019
		(0.047)	(0.042)		(0.036)	(0.033)
Risk taking		-0.075	-0.079**		-0.009	-0.010
		(0.047)	(0.036)		(0.041)	(0.033)
Trust		-0.009	0.015		-0.021	0.015
		(0.037)	(0.033)		(0.026)	(0.019)
Math		-0.010	-0.012		-0.004	-0.016
		(0.042)	(0.034)		(0.027)	(0.022)
Procrastination		0.066	0.036		-0.007	-0.015
		(0.053)	(0.047)		(0.047)	(0.037)
			Big-Five personality using Hahn et al. (2012)			
FiveC		0.082*	0.038		0.001	-0.019
		(0.044)	(0.042)		(0.036)	(0.028)
FiveA		-0.027	-0.041		0.008	-0.008
		(0.054)	(0.043)		(0.034)	(0.029)
FiveE		0.061*	0.064**		0.035	0.029
		(0.034)	(0.031)		(0.027)	(0.021)
FiveN		-0.023	0.010		-0.038	-0.015
		(0.030)	(0.025)		(0.027)	(0.023)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
FiveO		-0.053	-0.020		-0.021	0.001
		(0.048)	(0.043)		(0.027)	(0.025)
			Demographic characteristics			
Politics		0.068***	0.063***		0.018	0.005
		(0.023)	(0.022)		(0.016)	(0.017)
Economics		0.024	-0.005		0.039	-0.024
		(0.117)	(0.100)		(0.098)	(0.075)
Age		0.013*	0.015***		0.007	0.011**
		(0.006)	(0.005)		(0.006)	(0.005)
Female		0.117	0.132		0.139**	0.140***
		(0.085)	(0.081)		(0.064)	(0.053)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		0.214**	0.230***		-0.030	-0.047
		(0.098)	(0.076)		(0.072)	(0.058)
$>$ £ 800		0.200	0.205*		-0.071	-0.063
		(0.142)	(0.124)		(0.090)	(0.064)
			Ethnicity - baseline: Others			
Asian		0.193*	0.121		0.086	0.054
		(0.107)	(0.090)		(0.082)	(0.077)
White		0.089	0.060		0.023	0.044
		(0.114)	(0.091)		(0.096)	(0.079)
			Session - baseline: Session 1			
Session 2		-0.104	-0.230**		-0.058	-0.186***
		(0.102)	(0.090)		(0.078)	(0.070)
Session 3		-0.170*	-0.246***		-0.065	-0.102
		(0.098)	(0.093)		(0.081)	(0.085)
Session 4		0.016	-0.099		-0.063	-0.157*
		(0.104)	(0.082)		(0.095)	(0.083)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	856	832	832	856	832	832

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. DAMAGE: $\frac{R-S}{R}$, GAIN: $\frac{T-R}{R}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B35: Determinants of First-mover's Cooperation in the Charness-Rabin model: Symmetric games & CCEI > 0.80

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
RISK	-0.053	-0.065	-0.067
	(0.061)	(0.065)	(0.065)
TEMPT	-0.225**	-0.229**	-0.229**
	(0.100)	(0.097)	(0.094)
EFF	0.143**	0.128**	0.128**
	(0.057)	(0.053)	(0.052)
$\hat{\lambda}$			-0.058
			(0.119)
$\hat{\delta}$			0.285***
			(0.109)
CCEI			-0.135
			(0.454)
Round	-0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.002)
	the measures of Falk et al. (2018)		
P.Reciprocity		-0.007	-0.036
		(0.034)	(0.037)
N.Reciprocity		-0.043	-0.047
		(0.041)	(0.036)
Patience		-0.002	0.016
		(0.031)	(0.032)
Risk taking		0.064*	0.059**
		(0.036)	(0.024)
Trust		0.030	0.028
		(0.031)	(0.030)
Math		-0.095**	-0.086***
		(0.039)	(0.032)
Procrastination		-0.014	-0.029
		(0.042)	(0.032)
	Big-Five personality using Hahn et al. (2012)		
FiveC		-0.039	-0.038
		(0.033)	(0.025)
FiveA		-0.002	0.004
		(0.040)	(0.036)
FiveE		-0.086***	-0.100***
		(0.028)	(0.022)
FiveN		0.003	-0.012
		(0.024)	(0.020)

	CCEI > 0.80	CCEI > 0.80	CCEI > 0.80
FiveO		0.042	0.005
		(0.033)	(0.030)
	Demographic characteristics		
Politics		-0.043**	-0.045**
		(0.018)	(0.018)
Economics		0.016	0.044
		(0.090)	(0.085)
Age		-0.000	-0.002
		(0.005)	(0.005)
Female		-0.080	-0.144**
		(0.066)	(0.071)
	Expenditure per month, excluding rent - baseline: £ 400 to 800		
< £ 400		0.002	-0.027
		(0.076)	(0.066)
> £ 800		0.031	0.062
		(0.076)	(0.082)
	Ethnicity - baseline: Others		
Asian		0.174***	0.185***
		(0.061)	(0.047)
White		0.137**	0.164**
		(0.069)	(0.064)
	Session - baseline: Session 1		
Session 2		-0.013	0.053
		(0.072)	(0.060)
Session 3		0.144*	0.198**
		(0.079)	(0.077)
Session 4		-0.124	-0.017
		(0.094)	(0.071)
Control	*No*	*Yes*	*Yes*
Obs.	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Table 4.B36: Determinants of First-mover's Cooperation in the Charness-Rabin model: Asymmetric games & CCEI > 0.80

	Ahead	Ahead	Ahead	Behind	Behind	Behind
RISK	0.395	0.343	0.347	0.061	0.075*	0.074*
	(0.274)	(0.291)	(0.291)	(0.040)	(0.040)	(0.039)
TEMPT	-0.274	-0.332	-0.342	-0.224**	-0.227**	-0.228**
	(0.276)	(0.296)	(0.302)	(0.101)	(0.093)	(0.093)
EFF	0.526**	0.484*	0.490**	0.037	0.045	0.045
	(0.245)	(0.251)	(0.249)	(0.041)	(0.040)	(0.039)
$\hat{\lambda}$			-0.353***			-0.085
			(0.081)			(0.052)
$\hat{\delta}$			0.190**			0.020
			(0.077)			(0.062)
CCEI			0.407			0.233
			(0.484)			(0.402)
Round	-0.001	-0.001	-0.001	-0.002	-0.002*	-0.002*
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
			the measures of Falk et al. (2018)			
P.Reciprocity		0.044	0.029		0.027	0.022
		(0.035)	(0.028)		(0.018)	(0.018)
N.Reciprocity		-0.007	-0.015		-0.020	-0.016
		(0.040)	(0.032)		(0.018)	(0.019)
Patience		0.005	0.001		0.012	0.009
		(0.034)	(0.029)		(0.016)	(0.017)
Risk taking		0.035	0.033		-0.008	-0.006
		(0.037)	(0.029)		(0.013)	(0.014)
Trust		0.009	-0.021		0.009	0.006
		(0.024)	(0.020)		(0.012)	(0.013)
Math		-0.011	-0.011		-0.031*	-0.030*
		(0.028)	(0.024)		(0.018)	(0.017)
Procrastination		-0.015	-0.028		-0.040	-0.035
		(0.048)	(0.032)		(0.025)	(0.022)
			Big-Five personality using Hahn et al. (2012)			
FiveC		-0.054	-0.035		-0.030*	-0.018
		(0.036)	(0.027)		(0.017)	(0.014)
FiveA		0.024	0.034		-0.016	-0.012
		(0.031)	(0.027)		(0.018)	(0.017)
FiveE		-0.073***	-0.071***		-0.014	-0.016*
		(0.025)	(0.020)		(0.009)	(0.009)
FiveN		0.031	0.020		0.014	0.007
		(0.022)	(0.019)		(0.010)	(0.010)

	Ahead	Ahead	Ahead	Behind	Behind	Behind
FiveO		0.033	0.020		0.035**	0.030**
		(0.029)	(0.027)		(0.015)	(0.015)
			Demographic characteristics			
Politics		-0.010	0.001		-0.011	-0.008
		(0.017)	(0.014)		(0.010)	(0.010)
Economics		-0.049	-0.093		-0.033	-0.043
		(0.109)	(0.115)		(0.044)	(0.046)
Age		-0.010	-0.013***		-0.005*	-0.006**
		(0.007)	(0.004)		(0.003)	(0.003)
Female		0.076	0.031		-0.081***	-0.077**
		(0.064)	(0.044)		(0.029)	(0.030)
			Expenditure per month, excluding rent - baseline: £ 400 to 800			
$<$ £ 400		0.006	0.019		0.006	0.010
		(0.062)	(0.046)		(0.033)	(0.032)
$>$ £ 800		0.164	0.149		-0.019	-0.019
		(0.138)	(0.093)		(0.036)	(0.035)
			Ethnicity - baseline: Others			
Asian		-0.034	-0.033		-0.008	0.001
		(0.081)	(0.062)		(0.033)	(0.033)
White		-0.043	-0.083		0.028	0.017
		(0.085)	(0.064)		(0.040)	(0.039)
			Session - baseline: Session 1			
Session 2		0.086	0.171***		0.008	0.025
		(0.067)	(0.063)		(0.033)	(0.030)
Session 3		0.191**	0.219***		0.038	0.041
		(0.081)	(0.077)		(0.039)	(0.044)
Session 4		0.131	0.209***		-0.072	-0.044
		(0.087)	(0.066)		(0.044)	(0.034)
Control	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Obs.	688	664	664	688	664	664

Notes: Average marginal effects from random effect probit regression with robust standard errors clustered on individuals. RISK: $\frac{P-S}{P}$, TEMPT: $\frac{T-R}{T}$, EFF: $\frac{R-P}{R}$, P.Reciprocity: positive reciprocity, N.Reciprocity: negative reciprocity, Math (the self-assessment of the staement "I am good at math" on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly), Procrastination (the self-assessment of the staement "I tend to postpone tasks even if I know it would be better to do them right away." on an 11-point Likert scale, 0 means "does not describe me at all" and 10 means "describes me perfectly). FiveC (Conscientiousness), FiveA (Agreeableness), FiveE (Extroversion), FiveN (Neuroticism), and FiveO (Openness to Experience) are elicited from the Big-Five personality test of Hahn et al. (2012). Politics (0 to 10, 0 means the most left and 10 means the most right), Economics (1 if studies economics, otherwise 0) $^{***}p < 0.01$, $^{**}p < 0.05$, , $^{*}p < 0.10$.

Appendix 4.C Experimental Instructions

Welcome

Thank you for participating in Session 1 of this experiment. This experiment consists of two sessions: Session 2 will take place next week, and it will be entirely independent from Session 1.

Please refrain from the use of mobile phones during the experiment. Once the experiment begins, we ask everyone to remain silent. In order to keep your decisions private, please do not reveal your choices to any other participant. If you have any questions, please click "Ask the experimenter" button on the screen and an experimenter will approach your desk.

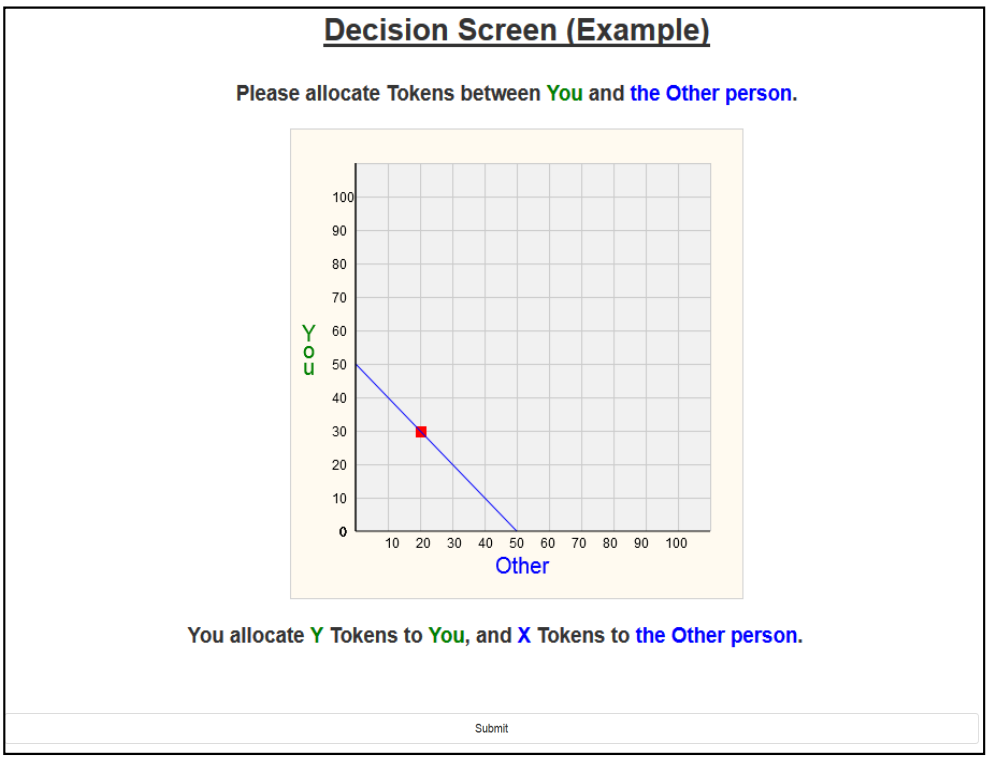
This is an experiment in decision-making. Your earnings will depend partly on your decisions and partly on the decisions of the other participants and partly on chance. Please pay attention to the instructions as a considerable amount of money is at stake.

Your participation in the experiment and any information about your earnings will be kept strictly confidential. Please note that your identity will not be revealed to anyone during the experiment. Thus, all your decisions will remain anonymous.

Details of how you will make decisions and receive earnings will be provided below.

Task

In this session, you will participate in 50 independent decision situations that share a common form. Each decision situation will be presented on a screen like the **example screen** below.



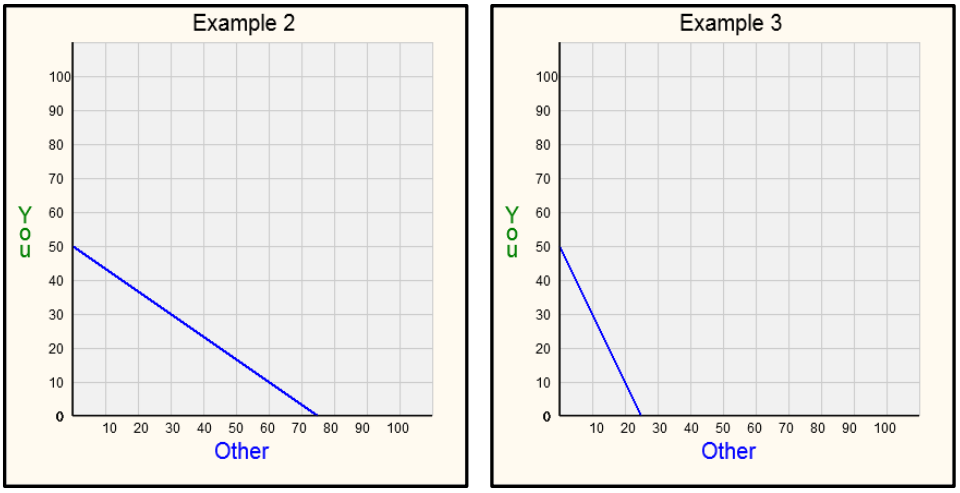
In each decision situation, you will be asked to allocate Tokens between **You** and **Other**. **Other** is one of the other participants in this room, randomly chosen in each decision situation. Each choice will involve choosing a point on a line representing possible Token allocations.

To choose an allocation, use the mouse: Click or drag to move the pointer on the computer screen to the allocation that you wish to choose. Once you use the mouse, you can also use the arrows on the keyboard to move the pointer. The computer will only allow you to choose allocations that are on the line.

After that, confirm your decision by clicking on the Submit button. Once you have clicked the Submit button, your decision cannot be revised.

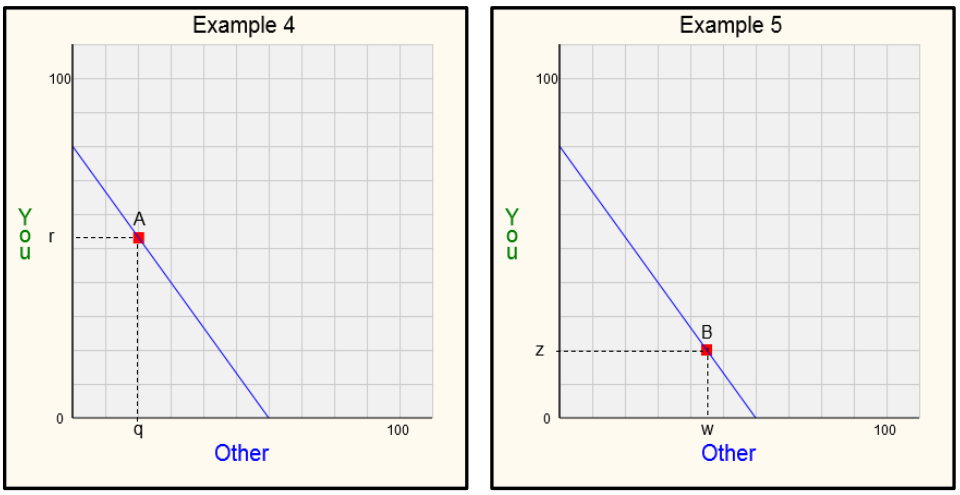
Next, you will be asked to make a decision in another independent situation. This process will be repeated until all 50 decision situations are completed. At that point, you may have to wait for other participants to finish.

Each decision situation will start by having the computer select a line randomly from the set of lines that (i) intersect with at least one of the axes at 50 or more Tokens and (ii) have no intercept above 100 Tokens or below 0 Tokens. Examples of lines you might face are shown in Examples 2 and 3.



If the line is **relatively flat** (Example 2), Other's Token increase by **more than 1 Token** as you decrease Your Token by 1 Token. If the line is **relatively steep** (Example 3), Other's Token increase by **less than 1 Token** as you decrease Your Token by 1 Token.

In each choice, you may choose any point that is on the line. For example, as illustrated in Example 4, choice A represents a decision to allocate r Tokens to You and q Tokens to Other. Similarly, in Example 5 choice B represents a decision to allocate z Tokens to You and w Tokens to Other.



Please note that in each decision situation,

1) the lines will differ and

2) your decision situation is independent of other participants' decision situations.

That is to say, the lines will differ for each participant in each decision situation. In the same way, other participant's decision situation will be independent of any participant's decision situation, and will not depend on your choices in any of the earlier decision situations.

During the Task, you will not receive any feedback on the other person's choice or the outcomes of the decision situations.

Your earnings

At the end of the session, you will be paid £ 3 as a participation fee. In addition, you will be paid additional earnings based on decisions at the end of Session 2 next week. You will collect your payment in cash and in private.

For your additional earnings in session 1, **one out of the 50 decision situations** will be randomly selected. To do this, one of you will be randomly chosen next week and he/she will draw a number from 1 to 50 in a jar. Your additional earnings will depend on the Tokens allocated in the corresponding decision situation. The Tokens will be converted to cash at the rate of 10 Tokens = £ 1.

For this randomly chosen decision situation, you will receive the Tokens you allocated to You. Another person who was matched with you in your decision situation will receive the Tokens you allocated to Other. In addition, there will be another participant in this experiment who was matched with you for the purpose of their own allocation decision. You will receive the Tokens this person allocated to Other. Please note this person is not the same person that you allocated Tokens to.

Welcome

Thank you for participating in Session 2 of the experiment. This session is entirely independent from Session 1.

Please refrain from the use of mobile phones during the experiment. Once the experiment begins, we ask everyone to remain silent. In order to keep your decisions private, please do not reveal your choices to any other participant. If you have any questions, please click “Ask the experimenter” button on the screen and an experimenter will approach your desk.

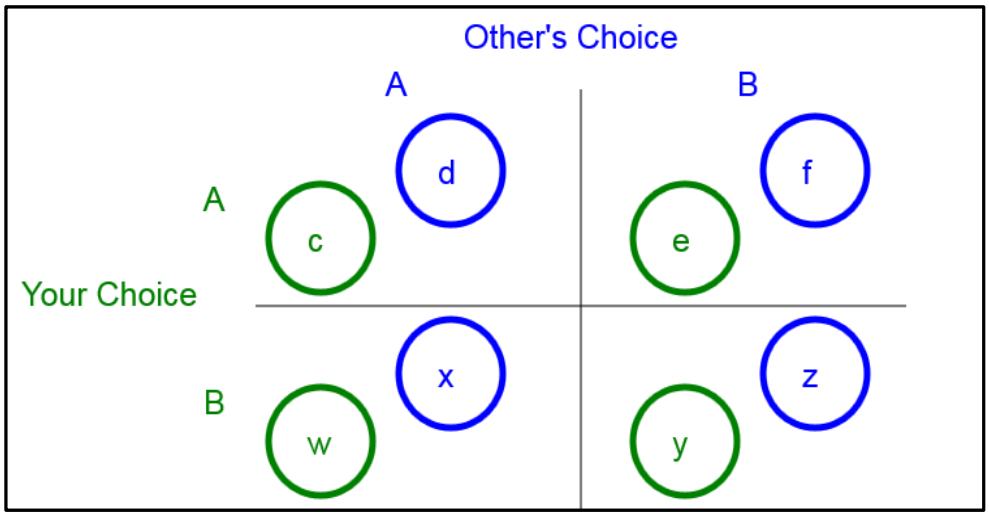
This is an experiment in decision-making. Your earnings will depend partly on your decisions and partly on the decisions of the other participants and partly on chance. Please pay attention to the instructions as a considerable amount of money is at stake.

Your participation in the experiment and any information about your earnings will be kept strictly confidential. Please note that your identity will not be revealed to anyone during the experiment. Thus, all your decisions will remain anonymous.

Details of how you will make decisions and receive earnings will be provided below.

Task

In this session, you will participate in 24 independent decision situations that share a common form. Each decision situation will be presented on a screen like the **example screen** below, except that all the letters will be replaced with numbers.



You and **Other** will be making choices between **A** and **B**. **Other** is one of the other participants in this room. **Your earnings** are the values in the green circle, and **Other's earnings** are the values in the blue circle. The table is read as follows:

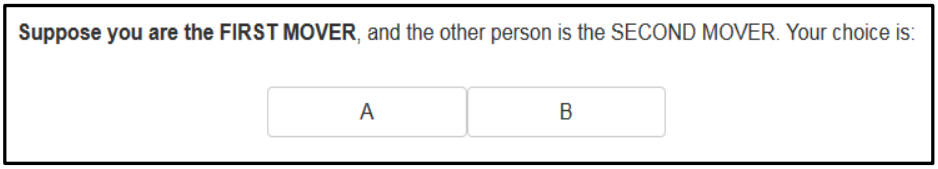
- If **You** choose A and **Other** chooses A, **You** will earn **c** Tokens and **Other** will earn **d** Tokens.
- If **You** choose A and **Other** chooses B, **You** will earn **e** Tokens and **Other** will earn **f** Tokens.
- If **You** choose B and **Other** chooses A, **You** will earn **w** Tokens and **Other** will earn **x** Tokens.
- If **You** choose B and **Other** chooses B, **You** will earn **y** Tokens and **Other** will earn **z** Tokens.

Please note that the values in the table will differ in each decision situation.

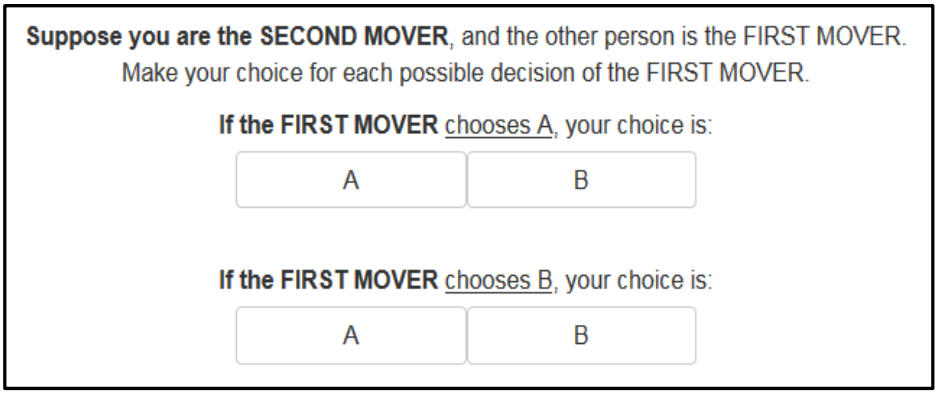
In each decision situation, you must complete **two types** of tasks, which we will refer to below as the "FIRST MOVER's decision" and "SECOND MOVER's decision". The FIRST MOVER decides first whether to choose A or B. The SECOND MOVER is then informed of the FIRST MOVER's decision. The SECOND MOVER then decides whether to choose A or B.

We want to know what you would do in the role of the FIRST MOVER and what you would you do in the role of the SECOND MOVER. Thus you will be prompted to make decisions in both roles.

● For the "FIRST MOVER's decision" task, you will see the following screen and you must choose A or B:



● For the "SECOND MOVER's decision" task, You will see the following screen and you must choose A or B in two possible cases: (1) if the FIRST MOVER chooses A (2) if the FIRST MOVER chooses B.



During the Task, you will not receive any feedback on the other person's choice or the outcomes of the decision situations.

Your earnings

At the end of the session, you will be paid £ 3 as a participation fee plus additional earnings based on decisions. You will collect your payment in cash and in private.

To determine your additional earnings, you will be matched with another randomly selected person in this room. Then, the computer will randomly **select one out of the 24 decision situations.** The computer will randomly choose either You or Other to be the FIRST MOVER.

If You are chosen as a FIRST MOVER, your FIRST MOVER's decision will be matched with the SECOND MOVER's decision of Other. If Other is chosen as a FIRST MOVER, your SECOND MOVER's decision will be matched with the FIRST MOVER's decision of Other. These decisions will determine your Token earnings and Other's Token earnings for the randomly chosen situation. The Tokens will be converted to cash at the rate of 100 Tokens = £ 1.

At the end of the session, you will be informed of your choices and earnings for the randomly chosen situation.

Chapter 5

Conclusion

This thesis has contributed to deepening our knowledge of how individuals behave in social dilemmas. All three studies are closely related to each other, and we especially focus on how payoff variations and preferences affect cooperation.

Chapter 2 examines the effect of payoff variations on cooperation in simultaneous-move prisoner's dilemmas. The indices of risk, temptation, and efficiency (Mengel, 2018) could plausibly affect cooperation. Previous studies do not provide clear evidence on which of these three factors is the most influential determinant of cooperation in that correlations between risk, temptation, and efficiency in previous experiments make it difficult to interpret which one of these three factors induces cooperation/defection. To address this limitation, we design a within-subject experiment where the levels of risk, temptation, and efficiency orthogonally vary across eight games.

We find evidence that the measure of temptation has a significantly negative impact on cooperation. Efficiency also has a positive effect on cooperation, but the effect size is small. Risk does not have a significant effect on cooperation. Our findings are robust to using other payoff indices (e.g., fear, greed, normalized loss and gain) that are relevant to risk and temptation.

Chapter 3 is the first study to examine the within-subject variability of conditional cooperation across payoff variations. We employ the payoff parameterization of chapter 2 but use *sequential* prisoner's dilemma games. Using a strategy method, we elicit second-mover strategies across eight games.

We find that conditional cooperation varies across games. Most subjects neither always conditionally cooperate nor always free-ride. There exist systematic patterns between conditional cooperation and payoffs: subjects conditionally cooperate more often when free-riding imposes larger losses on the opponent, or when free-riding provides smaller gains for themselves.

Chapter 3 makes two important contributions to the literature. First, we show that a majority of subjects change their second-mover strategy when payoffs change, thus classifications of individuals as "conditional cooperators" or "free-riders" should not be generalized to other games with different material payoffs. This also suggests that the effectiveness of leading-by-example in promoting cooperation may be limited when the additional gains from free-riding become greater or free-riding cause small monetary losses to the opponent. Second, our results support the view that conditional cooperation reflects underlying social preferences. The variation in patterns of conditional cooperation across payoffs are consistent with the predictions of several distributional preference models (e.g., Andreoni and Miller, 2002; Fehr and Schmidt, 1999).

Chapter 4 examines more directly the relationship between distributional preferences and conditional cooperation. Conditional cooperation can be accommodated by models of distributional preferences, which capture motives such as altruistic concern for another's earnings as well as their own earnings, a desire to equalize earnings, or a desire to maximize combined earnings. Although several studies (Blanco et al., 2011; Teyssier, 2012) find some evidence

that distributional preferences determine conditional cooperation, there are two limitations. First, these studies only investigated how inequality aversion affects conditional cooperation. Therefore, previous studies cannot examine how concerns for other's earnings and efficiency concerns, which also could plausibly explain conditional cooperation, affect conditional cooperation. Second, the predictive power of distributional preferences for conditional cooperation is not tested under a broad set of payoff parameterizations as these studies employed a one-shot game with fixed payoffs.

We use a CES utility function (Andreoni and Miller, 2002) that incorporates all of these motives. We design a within-subject experiment consisting of two parts a week apart: we estimate individual-level distributional preferences from fifty-round of modified dictator games (part 1), and we elicit first-mover choices and second-mover choices across twenty-four sequential prisoner's dilemma games where payoffs vary (part 2).

We find that many subjects' second-mover strategies are predicted by their distributional preferences. There is a clear positive relationship between fair-mindedness and conditional cooperation. Subjects who substantially care about other's earnings as well as their own earnings are more likely to conditionally cooperate but less likely to free-ride. We also find that distributional preferences are influential determinants of first-mover's cooperation. Our findings that distributional preferences predict conditional cooperation and first-mover's cooperation, are also valid when we use other distributional preference models (Charness and Rabin, 2002; Fehr and Schmidt, 1999).

Chapter 4 makes two main contributions to the literature on the role of social preferences in games. First, compared to previous studies, we provide stronger evidence that distributional preferences influence conditional coopera-

tion by testing the predictive power of distributional preferences under a broad set of payoff parameterizations in twenty-four games. Second, our finding also supports recent experimental studies (Ghidoni et al., 2019; Hedegaard et al., 2019; Miettinen et al., 2020) that report distributional preferences also play a role under strategic uncertainty. Especially, our results are in line with the discussion of Miettinen et al. (2020), which suggests that the first-mover's decision can be accommodated by social preferences and subjective beliefs about the second-mover's cooperation.

Future research could extend our focus from prisoner's dilemma games to public goods games. Existing studies (Fischbacher, Gächter and Fehr, 2001; Fischbacher and Gächter, 2010) report that many subjects who exhibit conditional cooperativeness have a self-serving bias in that they contribute less than others' average contribution. This pattern, known as imperfect conditional cooperation, can be accommodated by the distributional preference model of Andreoni and Miller (2002). Therefore, it would be interesting to examine how well distributional preferences predict imperfect conditional cooperation in public goods games.

Another potential future research area could be to examine more systematically different elicitation methods for eliciting distributional preferences to examine which method provides better predictors of subjects' behavior. A number of experimental methods for eliciting distributional preferences have been introduced (Blanco, Engelmann and Normann, 2011; Fisman, Kariv and Markovits., 2007; Kerschbamer, 2015; Kurschilgen, 2014; Yang, Onderstal and Schram, 2016), but it has not been examined so far which methods provide better predictive power.

References

Andreoni, J., and J. Miller. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, 70(2), 2002, 737–753.

Blanco, M., D. Engelmann, and H. T. Normann. "A Within-Subject Analysis of Other-Regarding Preferences." *Games and Economic Behavior*, 72, 2011, 321–338.

Charness, G., and M. Rabin. "Understanding Social Preferences with Simple Tests." *The Quarterly Journal of Economics*, 117(3), 2002, 817–869.

Fehr, E., and K. M. Schmidt. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114, 1999, 817–868.

Fischbacher, U., S. Gächter and E. Fehr. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, 71(3), 2001,

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

Fisman, R., S. Kariv, and D. Markovits. "Individual Preferences for Giving." *American Economic Review*, 97(5), 2007, 1858–1876.

Ghidoni, R., B. L. Cleave, and S. Suetens. "Perfect and Imperfect Stragners in Social Dilemmas." *European Economic Review*, 116, 2019, 148–159.

Hedegaard, M., Kerschbamer, R., Müller, D., Tyran, J.-R., "Distributional preferences explain individual behavior across games and time." SSRN 3388681, 2019.

Kerschbamer, R. "The geometry of distributional preferences and a non-parametric identification approach: The Equality Equivalence Test." *European Economic Review*, 76, 2015, 85-103.

Kurschilgen, M. "Identity Utility and Other-Regarding Preferences - an Experiment" Mimeo, Max Planck Institute for Research on Collective Goods, 2014.

Mengel, F. "Risk and Temptation: A Meta-Study on Prisoner's Dilemma Games." *The Economic Journal*, 128, 2018, 3182-3209.

Miettinen, T., M. Kosfeld, E. Fehr, and J.W. Weibull. "Revealed preferences in a sequential prisoners' dilemma: A horse-race between six utility functions." *Journal of Economic Behavior and Organization*, 173, 2020, 1-25.

Teyssier, S. "Inequity and risk aversion in sequential public good games." *Public Choice*, 151, 2012, 91-119.

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