Three Essays on Cooperation and Reciprocity

Ali Seyhun Saral
School of Social Sciences
University of Trento

A thesis submitted for the degree of
Doctor of Philosophy

Trento 2019

Three Essays on Cooperation and Reciprocity

Author Ali Seyhun Saral

Supervisors

Prof. Luciano Andreozzi, University of Trento

Prof. Matteo Ploner, University of Trento

Assesment Commitee

Prof. Topi Miettinen, Hanken School of Economics

Prof. Daniele Nosenzo, University of Nottingham

Examination Commitee

Prof. Marco Bee, University of Trento

Prof. Ericka Costa, University of Trento

Prof. Maria Vittoria Levati, University of Verona

Acknowledgements

Unfortunately, to list all the names that deserve to be here is basically an infeasible task. During my academic life I relied on the kindness of a vast number of people. Nevertheless, I will try to mention a small subset of them.

First of all, I would like to thank my supervisors Luciano Andreozzi and Matteo Ploner. I felt very privileged to have such great mentors. Besides countless things that I learned from them, I would also like to thank them for making hard-work enjoyable.

I would like to thank Topi Miettinen and Daniele Nosenzo for their evaluations and valuable comments which allowed me improve this dissertation substantially.

During my PhD, I benefited so many times from the discussions and conversations with the CEEL members. I would like to specifically mention Dominique Cappelletti, Marco Faillo, Luigi Mittone, and Marco Tecilla for their invaluable contributions. I am also truly grateful to Nicole Bertotti, Davide Santuari and the PhD office for their constant help.

Moreover, I would like to thank my wonderful colleagues and friends Bassim Allaheeb, Tatiana Balmus, Samuele Bibi, Ceren Battal, Sergiu Burlacu, Chiara Cali, Patricia Dragan, Austėja Kažemekaitytė, Chiara D'archangelo, Dimtri Kirov, Lena Kirova, Hamid Oskorouchi, Costanza Piovanelli, Christos Perentis, Piero Ronzani, Jonė Tamulaitytė, Ade Tett, Maria Tsouri, Nives Della Valle, Thanh Tam Vu, George Young, and Sally Young for making Trento what it is for me.

I would like to thank Ayça Ebru Giritligil for introducing me to the experimental economics and being a permanent mentor for me. I would also like to thank Valentina Carbonara for encouraging me to come to Trento. I am grateful to my friends Orhan Ayraç, Fatma

Aslan, Francesco Ciprian, and Pelin Yargıç for your support for the inspiration.

A significant amount of this dissertation has been written at the Max Planck Institute for Research on Collective Goods, Bonn. I would like to thank Matthias Sutter and Christoph Engel for giving me the opportunity and all the institute employees for their help. I would like to mention Zvonimir Bašić, Stefania Bortolotti, Brian Cooper, Zita Coppinger, Zwetelina Iliewa, Lukas Kiessling, Sofia Monteiro, Matthias Praxmarer, Shambhavi Priyam, Angelo Romano, Daniel Salicath, Sebastian Schneider, Sebastian Tonke, Eugenio Verrina, and Claudia Zoller for their help and support.

Special thanks to Piero; without whom, I wouldn't be able to survive the PhD for sure,

to Gabriela for her support and for being there for me,

and to my family: Göksu, Seyhan and Abdullah for all the sacrifices they did for me and for their long-standing hope that I would stop being a student at a certain point in my life.

Contents

List of Figures

Chapter 1

Introduction

Introduction

Perhaps one of the most compelling reasons to understand social behavior is to be able reach socially optimal outcomes that can only be acquired by mutual cooperation. How cooperation emerges, how it can be sustained, and in which conditions it is prone to collapse are some of the questions we ought to answer if we want to design better institutions. Although humans are exceptional in cooperating with unrelated individuals (Bowles and Gintis, 2011; Seabright, 2010; Nowak and Highfield, 2011), this ability should not be taken for granted.

Beside a wide variety of examples of the tragedy of the commons in real life, it is not uncommon either to see cooperative efforts fail. In most of these cases, failures of cooperation are due to the conflict between individual and social interests: cooperation is costly for the individual; yet, once achieved, its benefit can be enjoyed without paying a cost. As soon as a sufficient number individuals defect, the cooperation tends to collapse as a whole. Thus, studying motivations, preferences and incentives in a social context is essential to understand, and therefore to promote, social cooperation.

Economists have long been interested in cooperation, especially in the field of game theory. A paradigmatic example is the Prisoner's Dilemma, which demonstrates the conflict between individual and social interest in a simple framework. In economics, the Prisoner's Dilemma has been studied extensively since the 1950s, almost entirely with the assumption of self-interested individuals. However, during the last few decades, studies on social preferences in economics has been steadily growing, possibly as a result of developments in the experimental methods (Camerer, 2011). A rich stream of experimental results from several social dilemmas like the Ultimatum Game, the Public Goods Game, the Dictator Game, and the Trust Game has provided evidence that, contrary to standard economic theory, most individuals are socially-minded. They are willing to sacrifice their own payoffs to increase, or sometimes even to decrease, other people's payoffs. Soon after, behavioral models of reciprocity and inequity aversion emerged. Together with the insights from other disciplines like biology, psychology and political science, which share an interest in cooperation, those advancements illustrated new ways in which social cooperation could be sustained. While the early economic literature concentrated mostly on limited mechanisms to sustain cooperation such as reputation and repeated games effect, with the recent literature on social preferences a new range of possibilities emerged. If people's preferences deviated from pure altruism, cooperative outcomes could be sustained even in anonymous, non-repeated settings which are hostile to the emergence of cooperation.

This dissertation is a modest contribution to the literature of cooperation and social preferences. We use experimental and computational methods to understand the role and extent of reciprocity on cooperation.

Chapter 2 - *On the Stability of Conditional Cooperation* is a methodological contribution to the large literature on conditional cooperation in the Prisoner's Dilemma and the Public Good Game. In two seminal studies, Fischbacher et al. (2001) and Fischbacher and Gächter (2010) argued that most of the subjects involved in Public Good experiments can be classified as conditional cooperators. They are only ready to cooperate even in anonymous, non-repeated settings, only if they believe the others are doing the same. They stand ready to withdraw their contribution if they realize that other people are free-riding. These experiments have been replicated several times in a wide variety of contexts and in different cultures, and they show remarkably consistent results. However, some recent contributions have cast doubts on the robustness of these results. Burton-Chellew et al. (2015) and Burton-Chellew et al. (2016) show that reciprocal cooperators are more likely to be confused by the nature of the game, compared to the selfish subjects. These results suggest that there might be room for *learning* reciprocally cooperative preferences. In other words,the subjects' preferences may be prone to change as the game unfolds and experience accumulates. To test this hypothesis, we elicit conditionally cooperative preferences in an extended variant of sequential Prisoner's Dilemma and we let subjects play the game for ten rounds. Differing from the previous literature, we allow subjects to modify their conditional strategy, and not just their unconditional contribution, in each round. Our study replicates the result of previous studies in its first stage. More than half the subjects can be classified as conditional cooperators. Yet, as the game is played repeatedly, the fraction of conditionally cooperative subjects declines over time, while the fraction of selfish subjects increases. In the last round, the majority of subjects can be classified as purely selfish. The implications of our result is mostly methodological. In the current literature there, is a marked tendency to assume that the conditionally cooperative preferences revealed at the beginning of the experiment are subjects' "true" preferences. Our study draws a clearer picture as it points out that these preferences are prone to change during the course of the experiment.

Chapter 3 - *The Evolution of Conditional Cooperation* evaluates the evolutionary success of conditional preferences by using computer simulations. We use an agent-based model in which agents play a variation of the iterated Prisoner's Dilemma game. We consider a discrete version of reactive strategies (Wahl and Nowak, 1999a) that are aligned with the classification of conditional strategies introduced by classical studies in conditional cooperation in a Public goods game (Fischbacher et al., 2001; Fischbacher and Gächter, 2010). This allows us to connect the experimental literature on conditional preferences with the literature on theoretical and computational studies of the Prisoners' Dilemma. We estimate the likelihood of cooperation levels as well as the likelihood of the existence of conditional types for different continuation probabilities. We show that an all-or-none type of conditional cooperation strategy together with the perfect conditional cooperation strategy are most likely to emerge when the continuation probability is sufficiently high. Our most surprising finding is related to the so-called hump-shaped strategy, a conditional type that is commonly observed in experiments. Our simulations show that those types are likely to thrive for intermediate levels of the continuation probability due to their relative advantage when probability of interaction is not enough to sustain a full-cooperation, but instead merely sustains mid-level cooperation.

Finally, Chapter 4 - *Presumptive Reciprocity in Dictator Games* aims to understand the underlying reciprocal motives in altruistic behavior. We argue that the altruism that is revealed in dictator games can be explained by what we call *presumptive reciprocity.* Subjects may display non-selfish preferences because they *presume* that the other subjects would have revealed similar, non-selfish preferences if the roles had been reversed. This kind of intuitive reasoning, although partially captured by indirect reciprocity, is overlooked in the literature on social preferences, especially when it comes to explaining the behavior that appears to be purely altruistic. The experimental evidence we provide shows that people's choices reveal mostly presumptive reciprocity, while purely altruistic preferences play a much smaller role.

Chapter 2

On the Stability of Conditional Cooperation

with Luciano Andreozzi and Matteo Ploner

Abstract

An often-replicated result in the literature on social dilemmas is that a large share of subjects reveal conditionally cooperative preferences. Cooperation generated by this type of preferences is notoriously unstable, as individuals reduce their contributions to the public good in reaction to other subjects' free-riding. This has led to the widely-shared conclusion that cooperation observed in experiments (and its collapse) is mostly driven by imperfect reciprocity. In this study, we explore the possibility that reciprocally cooperative preferences may themselves be unstable. We do so by observing the evolution of subjects' preferences in an anonymously repeated social dilemma. Our unsettling result is that, in the course of the experiment, a significant fraction of reciprocally cooperative subjects become egoistic, while the reverse is rarely observed. The non-selfish preferences that appear to be more stable are those most easily attributed to confusion. We are thus driven to the conclusion that egoism is more resistant to exposure to social dilemmas than reciprocity.

Keywords: altruism | strategy method | reciprocity | social preferences

2.1 Introduction

The large experimental literature on experiments on social dilemmas like the Public Goods Game, the Prisoner's Dilemma and the Trust Game has been remarkably coherent in revealing two stylized facts. First, subjects cooperate more than they would do if they were rational and purely egoistic. Second, when the game is repeated anonymously, cooperation declines over time, although it rarely disappears altogether (Ledyard, 1994; Chaudhuri, 2011; Cooper and Kagel, 2011). Engle-Warnick and Slonim (2006) document the decline of cooperation in the Trust Game, even when repeated between the same subjects). At the cost of a somewhat drastic simplification, the early explanations for this phenomenon can be grouped under two headings: the *learning* and the *reciprocity* hypothesis (see for example, Cooper and Stockman (2002) and Burton-Chellew et al. (2015)). According to the learning hypothesis, subjects are mostly self-interested, but they are boundedly rational and it takes time for them to learn the working of the game they are playing. For example, they may have trouble understanding the logic of dominant strategies or backward induction. According to this view, what is observed in the early stages of any experiment involving a social dilemma is a mixture of noise and confusion (Gale et al., 1995; Palfrey and Prisbrey, 1997).

This explanation has been repeatedly found wanting. In a classic article, Andreoni (1995) concluded that bounded rationality and learning could explain no more than half of the observed deviations from self-interest. His verdict was that "cooperation often observed in [...] public good experiments may not be due to learning, but instead may be due to frustrated attempts at kindness" (892). The obvious alternative to the learning hypothesis was that the cooperation observed in the lab (and its decline) could be the consequence of some form of pro-social behavior mostly driven by reciprocal altruism. This view is intuitively appealing. Reciprocally altruistic preferences were documented in other fields like social psychology (Kelley and Stahelski, 1970) and were considered a more natural alternative to straight selfishness by economists and philosophers alike (Sen, 1996; Sugden, 1984; Rabin, 1993). They had also some indirect empirical support. Deviations from pure self-interest were observed in simpler games, like the sequential Prisoner's Dilemma (Clark and Sefton, 2001a) and even in decision settings like the Dictator Game (Andreoni and Miller (2002) and Fisman et al. (2007)), in which they were harder to attribute to confusion.

However, the most compelling evidence against the learning hypothesis came directly from Public Goods Game experiments. In two widely quoted papers, Fischbacher et al. (2001) and Fischbacher and Gächter (2010) used the strategy method to elicit subjects' preferences for conditionally altruistic behavior. Their data revealed that around a half of the subjects could be classified as reciprocally cooperative: they were willing to contribute to the public good only if other subjects were contributing as well. Unconditional defectors where slightly more than 20 percent.

Cooperation sustained by reciprocally altruistic preferences is inherently fragile. While most subjects are willing to give less than what they expect others to give, none is willing to give more. So cooperation would unravel over time, even in the absence of free-riders (Fischbacher and Gächter, 2010, p. 554). These experimental findings proved to be robust (Kurzban and Houser, 2005; Kocher et al., 2008; Neugebauer et al., 2009; Herrmann and Thöni, 2009), and the reciprocity hypothesis became the dominant explanation in the experimental literature. For example, after a careful reading of the whole literature, Chaudhuri (2011) concludes that the "most notable finding [...] is that many participants behave as "conditional cooperators", whose contribution to the public good is positively correlated with their beliefs about the contributions to be made by their group members" (49).[1]

Despite its success, the reciprocity hypothesis had a weak spot. One of its implicit assumptions is that the reciprocally cooperative preferences revealed at the beginning of the experiment are the subjects' "true" preferences. However, this begs the original question, because proposers of the learning hypothesis may retort that *all* preferences are subject to change as the game unfolds and experience accumulates. The conditionally cooperative preferences revealed at the beginning of the experiment should not be an exception.

We should then be open to the possibility that at least a share of the reciprocity observed in experiments may be due to an imperfect understanding of the nature of the game that is being played. Indeed, there are reasons to believe that this may be the case. It is a long-known fact that a fraction of the subjects involved in social dilemmas display the same pro-social behavior in interactions with computers and human beings alike (Houser and Kurzban, 2002). Burton-Chellew et al. (2016) have recently shown that conditionally cooperative subjects are precisely those who seem unable to distinguish computers from human beings. They are also the only subjects who fail to answer control questions which aim to check their real understanding of the game. The authors conclude that there is "no evidence that there is a sub-population of players that understand the game and have pro-social motives".

This suggests that confused subjects may be of two different kinds. Some are of the familiar type, whose conditional contribution follows a non-discernible pattern. Others may display reciprocally cooperative preferences that, however, are determined by a faulty understanding of the game. For example, some subjects may fail to see that they are playing a repeated

[1] The learning hypothesis did not disappear altogether. For example, Arifovic and Ledyard (2012) suggest that the decay of cooperation can best be explained by models of learning, complemented with a hypothesis about the pro-social preferences of the individuals.

game which is anonymous, so non-cooperative choices will not have negative consequences on future interactions. This kind of mistake is clearly more difficult to detect, but it may be much more pervasive. To see why, consider the frequently-made observation that the choices made by inexperienced subjects reflect their day-by-day experience with situations they perceive as similar to the ones they encounter in the lab.[2] In a contribution that has received much attention, Rand et al. (2012) articulate a variant of this theory and dub it the *Social Heuristic Hypothesis* (SHH). According to the SHH, "people internalize strategies that are typically advantageous and successful in their daily social interactions. They then bring these automatic, intuitive responses with them into atypical social situations, such as most laboratory experiments. More reflective, deliberative processes may then override these generalized automatic responses, causing subjects to shift their behavior towards the behavior that is most advantageous in context."[3] The relevance of this approach for the theory of pro-social behavior stems from the fact that, in everyday life, repeated interactions are rarely anonymous, and in such settings neither unconditional defection nor unconditional cooperation are good strategies. (Axelrod, 2009; Nowak, 2006). Hence, it is just natural to expect that, in dealing with unfamiliar situations that are repeated, but anonymous, a non-negligible part of the subjects initially reveal reciprocally cooperative preferences.

The considerations just made suggest that the existing evidence on the Public Goods Game is stacked in favor of the reciprocity hypothesis. Since in ordinary life it usually pays to be reciprocally cooperative, one should

[2]Among others, this argument is used by Henrich et al. (2005) to explain the large variance in behavior observed in playing several games across culturally diverse societies, and by Gächter and Herrmann (2009) to explain the prevalence of anti-social punishment in societies with little social capital.

[3]An empirically verifiable consequence of the SHH is that, when choosing under time pressure (and hence forced to use their "intuition"), subjects will play more cooperatively than when they are given more time to reflect. The evidence collected by Rand et al. (2012) and Nishi et al. (2017) goes in this direction, although some of their results proved hard to reproduce (Tinghog et al., 2013).

expect this to be the instinctive, immediate choice in the early stages of any experiment involving social dilemmas. However, standard experiments in which reciprocal preferences are only elicited in the early phases of the sessions provide no information about their evolution as experience accumulates. More information could be obtained if subjects were to reveal their conditional choices along the entire duration of the game. This is where our contribution lies.

We consider a three-strategy sequential Prisoner's Dilemma. In a baseline setting we call *CondInfo*, we use the strategy method to elicit subjects' preferences as first and second-movers. When playing as second-movers, subjects have to state which level of cooperation they would choose in response to any of the three levels of cooperation the first mover may choose. After choices are made, roles are randomly assigned, first and second movers are informed about each other's respective strategy, and the payoffs are obtained. Notice that after the game is played each subject is informed about the way in which the other subject has filled the strategy method questionnaire. For example, he learns whether the subject he interacted with was an unconditional defector, a reciprocal cooperator or a confused subject. We run this for a total of ten periods. Differing from the existing experiments of the same kind, we allow subjects to express their choices as first movers *and* their conditional choices as second movers.

Our results can be summarized as follows:

- In the initial period, subjects' conditional preferences are no different from those observed in the many experiments involving Public Goods Games. The only slight difference is the lower incidence of confused players, which is probably due to our particularly simple setting. Sixty per cent of the subjects are evenly split between perfect reciprocators and purely egoistic types. The remaining forty per cent is made up of imperfect reciprocators. There is a residual fraction of confused subjects.

- In the subsequent periods, we observe a steady increase of purely egoistic types. By the last period, they total almost 60 per cent of the subjects.
- This increase is mostly due to subjects who are perfectly conditionally cooperative becoming unconditionally selfish over time. The fraction of subjects playing other strategy types, including those who are imperfect reciprocators, remains virtually unchanged.

It is tempting to interpret these results as a consequence of imitation and conformism (Bardsley and Sausgruber, 2005), but this would be misleading. In the first period, reciprocal cooperators are as abundant as perfect egoists and it is not clear why the firsts imitate the seconds and not vice versa. By the same token, it is hard to see why imitation is irrelevant for players who are imperfect cooperators or appear to be confused.

There are two more plausible explanations. First, reciprocal cooperators may be inclined to reciprocate not only the *actions* chosen by their partners when playing as first movers, but also the type of conditionally cooperative preferences they reveal when playing as second movers. This is a subtle, but crucial difference. Upon learning that the first mover has defected, a reciprocator playing as second mover will defect. This is the standard consequence of reciprocity which is captured by the strategy method and is deemed to be responsible for the decay of cooperation. In our experimental design, however, the subjects are also given information about the other subjects' conditional preferences as second movers. On learning that many of the other subjects are unconditional defectors (i.e., they would respond with defection to cooperation), a reciprocator may be induced to switch to unconditional defection as well. We shall call this type of reaction level-two reciprocity, in contrast to reciprocity based on the actions of the opponent, level-one reciprocity. The decay in the number of conditional cooperators can obviously be due to level-two reciprocity.

A second explanation is that preference instability is mostly due to a combination of selfishness and learning. When the interaction is anonymous, unconditional defection is the optimal strategy, as a reciprocally cooperative subject fails to exploit the other subjects who play cooperatively as first movers. Selfish subjects who learn from experience will soon discover that when the setting is anonymous there is no point in reciprocating the co-operative behavior of the first mover, and they will eventually become unconditional defectors.

To have a better understanding on the extent of the two explanations, we ran a second treatment, which we call *NoCondInfo*, in which subjects are given no information about the other subjects' conditional choices. If the decline in the number of reciprocal cooperators is mostly due to level-two reciprocity, then we might expect a sharper decay in the fraction of reciprocal cooperators in the setting in which information about the other subjects' conditional choices is provided. If instead the decay is mostly driven by learning, then we should expect similar patterns in two treatments. Indeed, our result in the second treatment is no different from the first. This result is in favor of the learning hypothesis, although we cannot discard that a certain degree of learning happens in both treatments. The possible implications of our results, and some possible avenues for future research, will be discussed in the closing section.

The question concerning the stability of conditionally cooperative preferences had been acknowledged in the very first contributions to this literature, but received comparatively little attention in the years that followed. Fischbacher et al. (2001) were the first to notice that their explanation for the decline of cooperation crucially hinged on the assumption that the preferences elicited with the strategy method remained stable, while the subjects' contributions declined. In Fischbacher and Gächter (2010), this problem is tackled by running two treatments of a ten-period Public Goods Game, in which subjects' conditional preferences are elicited either before or after the game is played. They found no difference between these treatments. Kurzban and Houser (2005) address the same issue using a different method. They first classify subjects with a standard Public Goods Game and then let them play three more social dilemmas. Their data show that the type of preferences subjects reveal in the first part of the experiment predicts the way they play in the second part. They conclude that their results "provide evidence that types [...] are different from one another and stable over time.". (p. 1805). The closest contribution to the present research is Volk et al. (2012), who investigate the stability of conditionally cooperative preferences in a longer time frame. They elicit the subjects' preferences twice: first on the day of the experiment and then in another session, either a half a month or five months later. Their data reveal that preferences are remarkably stable over time.

At this stage, we can only speculate about the reasons for the stark difference between the results we obtain and the ones present in the literature. To begin with, our experiment allows subjects to change their strategies repeatedly along the game, which allows for trial and error. Also, choices are revised right after payoffs are obtained, which presumably makes learning more effective. Finally, in the *CondInfo* treatment, we provide subjects with feedback about the other subjects' conditional choices, although, as we said in the Introduction, this seems to have a limited effect on the final outcome. Needless to say, more research is needed to disentangle the effects of learning and reciprocity in the evolution of the observed preferences.

2.2 Experimental Design

2.2.1 The Game

We used a three-strategy variant of a sequential Prisoner's Dilemma. Each player receives 100 tokens and is given the opportunity to transfer nothing (low transfer, L), 50 tokens (medium transfer, M), or 100 tokens (high transfer, H) to the other player. The second player chooses after having observed the first player's choice. A player's final payoff is the sum of the tokens he did not transfer, plus the tokens he received multiplied by three. The Pareto-optimal choice is thus to transfer 100 tokens, although the dominant choice is to transfer nothing. The extensive form of the game is represented in Figure 2.1. We will refer to this game as the Three Strategies Sequential Prisoner's Dilemma (3SSPD).

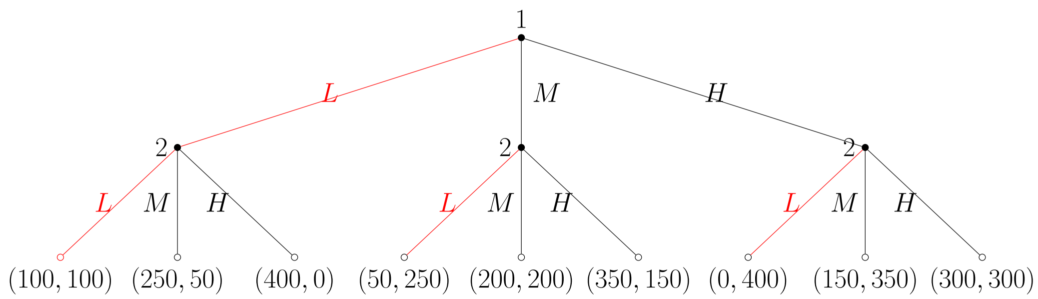

Figure 2.1: The Three-Strategies Sequential Prisoners Dilemma (3SSPD)

Although at first sight it may look otherwise, this game is a two-player variant of the Public Goods Game which is used by Fischbacher and Gächter (2010). To see this, consider a four-player Public Good Game in which conditionally cooperative preferences are elicited. This requires each subject to decide, before the game is played, how much to contribute to the public good (unconditional decision) and how much to contribute conditionally on the average contributions of the other players (conditional decision). To make the elicitation process incentive-compatible, it is necessary to impose a rule whereby the contribution of three of the four members of the group will be their unconditional contribution, while the fourth member's choice will be based on his conditional contribution. This amounts to assuming that the game has a sequential structure in which one of the players will be able to decide his contribution after having observed the other subjects' choices. The 3SSPD is just a version of this game with only two players.

The main reason for choosing this game, rather than the better-known Public Goods Game or the simultaneous Prisoner's Dilemma, is that we wanted to strike a balance between two competing necessities. On the one hand, we wanted a game rich enough to discriminate subjects who have reciprocally cooperative preferences from other types that are frequently observed in experiments of this kind. The simpler two-strategy Prisoner's Dilemma, for example, would not allow us to discriminate conditionally cooperative subjects from the so-called "hump-shaped" types, whose contribution is minimal for the low and high levels of the other subjects' contributions, and high for intermediate values. On the other hand, we wanted the game to be simple enough to avoid confusion. The main concern was that we wanted to elicit subjects' conditional choices several times during the course of the experiment. The conditional contribution table designed by Fischbacher et al. (2001), for example, requires a subject to make 21 distinct choices. It follows that, if the game is played repeatedly for ten periods, a subject has to enter her desired contribution levels up to 210 times during the experiment, in contrast to our setup, which requests 30 decisions for the conditional choices. Moreover, we believe that a sequential two-person game lends itself more naturally to elicit the conditionally cooperative preferences of the second mover (Andreoni and Samuelson, 2006). A key advantage is that it does not require players to understand the logic of dominant strategies (as in simultaneous Public Goods

Game) or of backward induction (as in the Ultimatum Game and any other game that has Nash equilibria that are not subgame-perfect.)

An issue on multiple-player games with conditional preference elicitation is that it is often unclear to what exactly subjects are conditioning their contribution (for instance, the mean contribution, the median contribution, the minimum contribution in the group, and so on.). A lack of information on the individual contribution in a Public Goods Game with groups leaves a lot of room for one participant to speculate about another participant's contribution. Moreover, even if two decision-makers have exactly the same preferences, their line of reasoning might differ. Hartig et al. (2015) shows that when information about individual contributions in a group are given a in Public Goods Game, contributions are higher comparing to the case which only group average is visible.

Finally, our setup creates a parameter-free classification of conditional cooperators. The common approach to detect conditional types has been to measure statistically the correlations between one's own contributions and the hypothethical contribution of the others they are conditioned to. This might be an issue, as borderline cases such as conditional preferences with a strong selfish bias might be classified differently due to small changes in the contribution schedule. Since we have three levels, each type of contribution pattern clearly fits to a type. Misidentification of types due to the noise in choices are unlikely.

2.2.2 Treatments

In all treatments, we used the strategy method to elicit the subjects' strategies as both first and second movers. Subjects are informed that they will be selected in either role with equal probability. When choosing their strategy as the second mover, each subject has to fill in a questionnaire in which he states his preferred strategy, conditional upon the strategy chosen by the first mover. Subjects played the 3SSPD game for 10 periods. To minimize the effect of strategic considerations, we follow the standard practice of using random protocol matching. In each period, subjects are paired randomly to play the game. and expressed their strategies both as first and as second movers. Before each period, we elicited subject's beliefs on their counterpart's choice. In order to have incentive compatibility, we awarded correct guesses by using the Quadratic Scoring Rule (Brier, 1950b), with the presentation procedure by Artinger et al. (2010). Additional information on belief elicitation can be found in Section A.2.

In order to evaluate the effect of extensive information on the conditional types of the opponent, we presented subjects with the following information scheme: in one treatment (NoCondInfo), after each period subjects were informed about whether they were the first or the second player, the action they played in the assigned role, the action chosen by their counterpart, and their payoff. In the second treatment (CondInfo), we also informed subjects whose roles had been determined as first movers about the conditional strategy chosen by their counterpart who played as the second mover. Every other step is identical for the two treatments.

2.2.3 Experimental Flow

Before the experiment, subjects were required to answer questions aimed at ascertaining their comprehension of the 3SSPD game. Payoffs were created randomly by the computer. Subjects could not proceed to the next step if they failed to provide correct answers to all questions. After completing this stage, four training periods took place, in which subjects played against randomly responding computer players. Following that, subjects were trained on the belief elicitation procedure. At this point the actual experiment took place: subjects played ten periods of the 3SSPD game against each other. Before closing the experiment, we gave them a short questionnaire which contained demographic questions. Table 2.1 summarizes the experimental flow.

Stage	**Repetition**
Control Questions	once
Training - Game	2 times
Training - Belief Elicitation Training - Game	2 times
Belief Elicitation Game	10 times
Questionnaire	once

Table 2.1: Sequence of Stages of the Experiment

Experimental sessions were conducted in CEEL, University of Trento. In total, 134 subjects participated in six experimental sessions. A copy of the instructions that were handed out in printed form is available in Appendix C.5. All subjects were able to answer the control questions correctly. No subject or session has been excluded from the data.

2.3 Results

2.3.1 Overview

Our first period results is in line with the previous experiments in repeated social dilemmas which we summarized in the Introduction: in the first period, the majority of subjects were revealed to be reciprocal cooperators and cooperation declines over time. We shall briefly discuss these findings in this order.

Most subjects are initially reciprocally cooperative. We represent a strategy for a subject as a triple ABC, where A is the action chosen in response to L, B in response to M, and C in response to H. For example, LMH is the perfectly reciprocating strategy that always matches the first mover choice. Given our simple strategic setting, the classification of subjects is immediate. We distinguish *selfish* individuals (LLL) from *perfect reciprocators* (LMH) and *imperfect reciprocators*, that is, all the subjects whose contribution is a monotonic function of the first-mover contribution (LMM, LHH, LLM, LLH). In our sample, there is a small group of *hump-shaped* subjects, whose contribution is higher in response to intermediate contributions by the other subjects (LHM or LML). All other subjects are labeled as "other patterns" [4]. Figure 2.2 represents the number of each type in the initial composition of the population in comparison with other studies of the same type. Moreover Table A.2 shows that the distribution of the conditional types are not significantly different from the distribution of those in Fischbacher et al. (2001) in the first eight periods while they are significantly different in the last two periods.

[4] A recent study by Fallucchi et al. (2018) classifies conditional strategies by using hierarchial clustering in the following way: own maximizers (OWN), strong conditional cooperators (SCC), weak conditional cooperators (WCC) and various (VAR). We also make a similar distinction of perfect (strong) and imperfect (weak) conditional cooperators. However the simple strategy space we have requires no such method to classify subjects' decisions in 3SSPD game, types staistically or computationally.

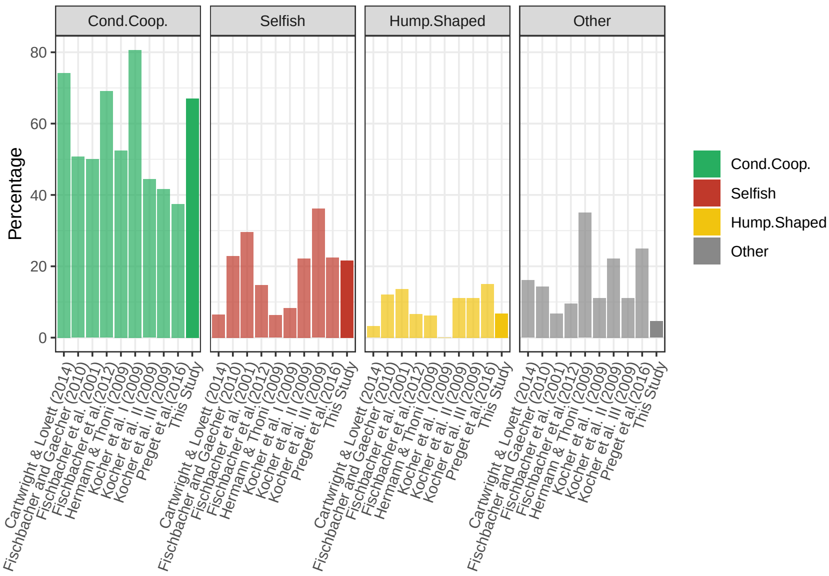

Figure 2.2: Comparison of Conditional Strategies on Period 1 with the Previous Studies.

Cooperation declines over time. Figure 2.3 describes the decline of cooperation over time, pooling together the results of the two treatments. An interesting pattern is that the fraction of subjects who choose the intermediate transfer M remains fairly stable over time both for the first and the second mover in response to H. Most of the observed decline in cooperation is thus explained by a shift from high to low transfers. Figure 2.4 represents the corresponding decay of the average payoff.

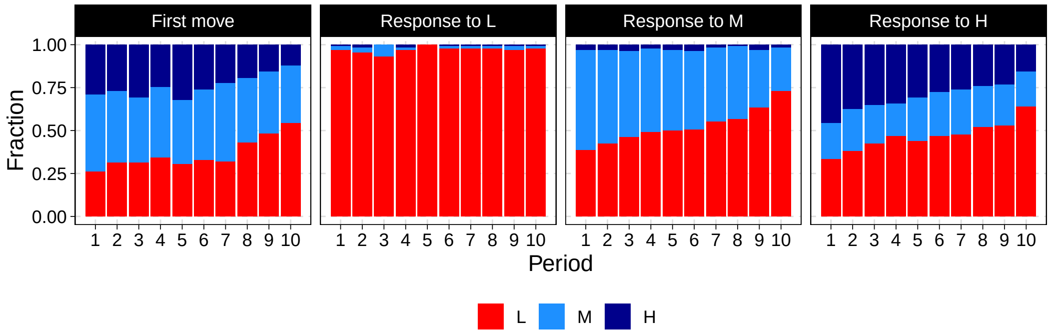

Figure 2.3: The Decline of Cooperation Over Ten Periods: Actions

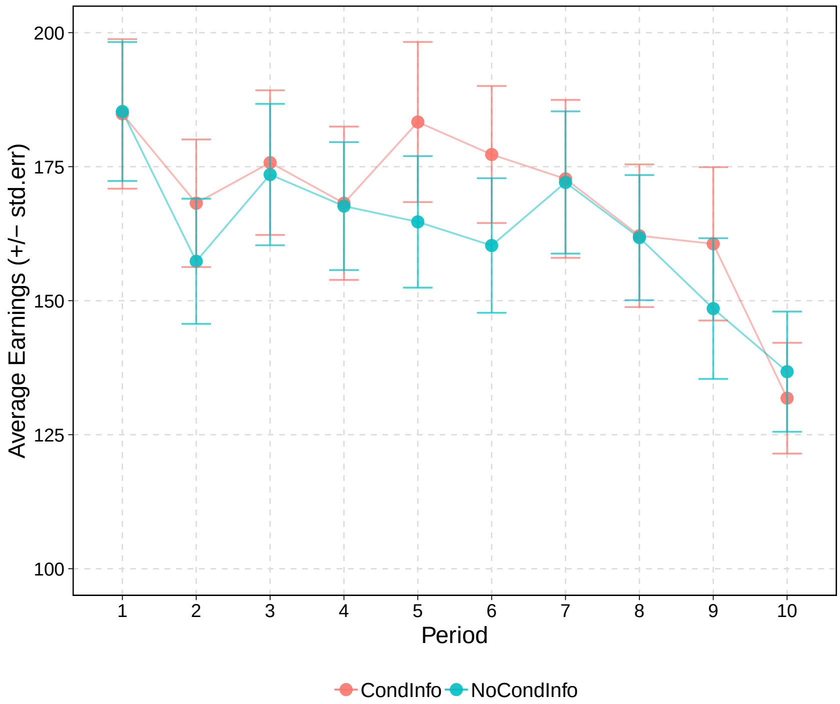

Figure 2.4: The Decline of Cooperation Over Ten Periods: Payoffs

2.3.2 Main Findings

If the conditional choices were collected only in the first period, one would be naturally led to the conclusion that the decline of cooperation was due to reciprocity. However, eliciting conditional prefernces in each period provides a more nuanced picture.

Result 1 - Pro-social preferences are unstable: Figure 2.5 represents the evolution over time of the choice made by each subject as second mover in both treatments. The final distribution among the strategies is very different from the initial one. In the final composition, the fraction of reciprocal cooperators (either perfect or imperfect) is down from 65% (62% for *CondInfo* and 68% in *NoCondInfo*) to 34% (36% in *CondInfo* and 32% in *NoCondInfo*. At the same time, the fraction of selfish individuals rises from 26% (29% for *CondInfo* and 24% in *NoCondInfo*) to 59 (56% for *CondInfo* and 62% in *NoCondInfo*). As Table A.3 and Table A.4 demonstrate, we found no significant differences between treatments in terms of fraction of cooperators or selfish subjects.

The red and dark green rectangles represent choices of subjects who are perfectly selfish (LLL) and perfect reciprocators (LMH). The light green rectangles pool together all imperfectly cooperative choices, while the yellow rectangles correspond to hump-shaped choices. All other patterns are represented by pink rectangles. This picture reveals that only a small minority of the subjects keep the same strategy throughout the game (26%). Some of them repeatedly switch between several strategies, while others switch only once.

Figure 2.5 suggests that subjects who are more prone to switch to selfishness are those who are initially classified as perfect reciprocators. This impression turns out to be true. Figure 2.6 shows the evolution over time of the number of strategies played by second movers in the two treatments. The change in the composition of the population is mostly due to the decline of perfect reciprocators and the increase in the number of selfish individuals. The frequency of all other strategies remains fairly constant across the periods.

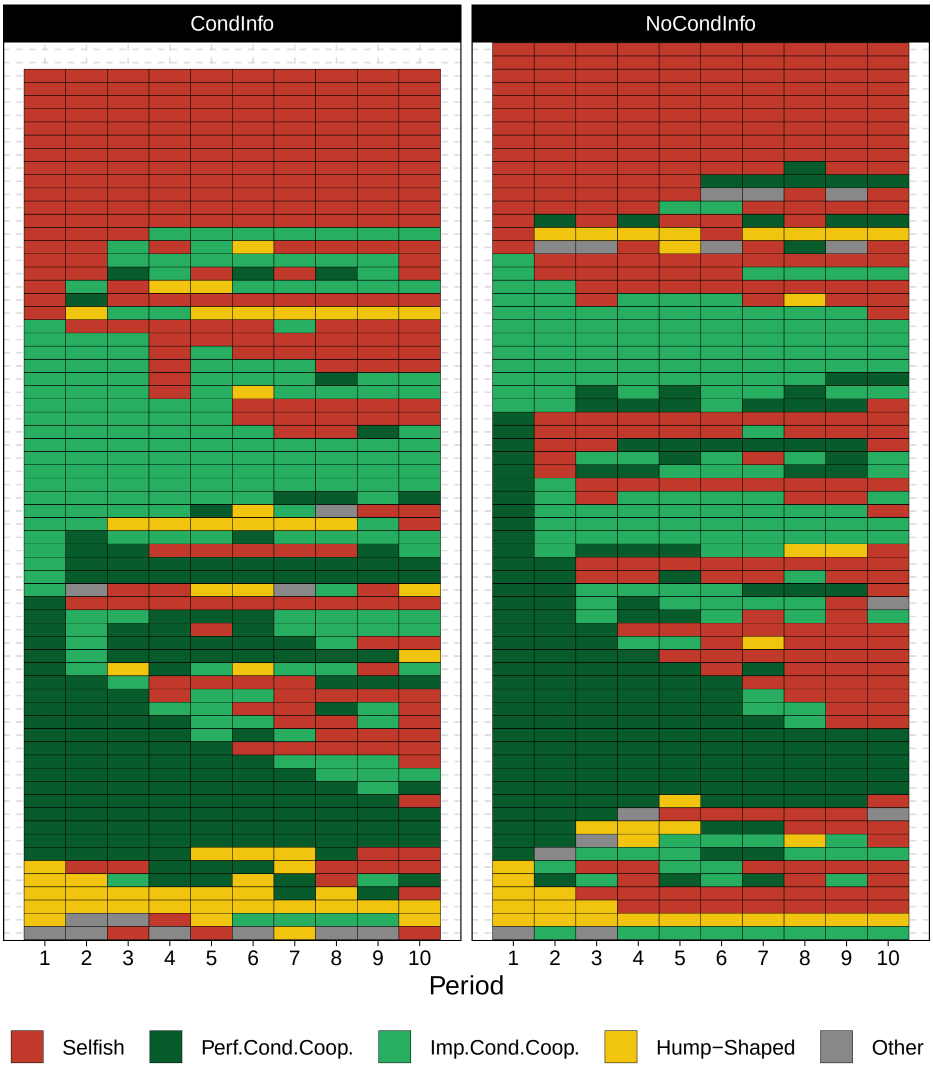

Figure 2.5: Subjects' Conditional Strategies Over Ten Periods for Each Treatment. In the plot, every row represent a subject and every rectangle with a particular color represent their conditional strategy in the period denioted in x axis.

Result 2 - Information has a limited effect on conditional strategies : A comparison between the treatments *NoCondInfo* and *CondInfo* reveals that it makes no difference whether one reveals the second mover's conditional choice. Table 2.2 shows results of a mixed effects logistics regression where the dependent variable is the dummy for using a conditionally cooperative strategy. The results of the analysis indicate that the additional information on the counterpart's conditional strategy does not have a significant effect on the choice of playing a conditionally cooperative strategy or not.

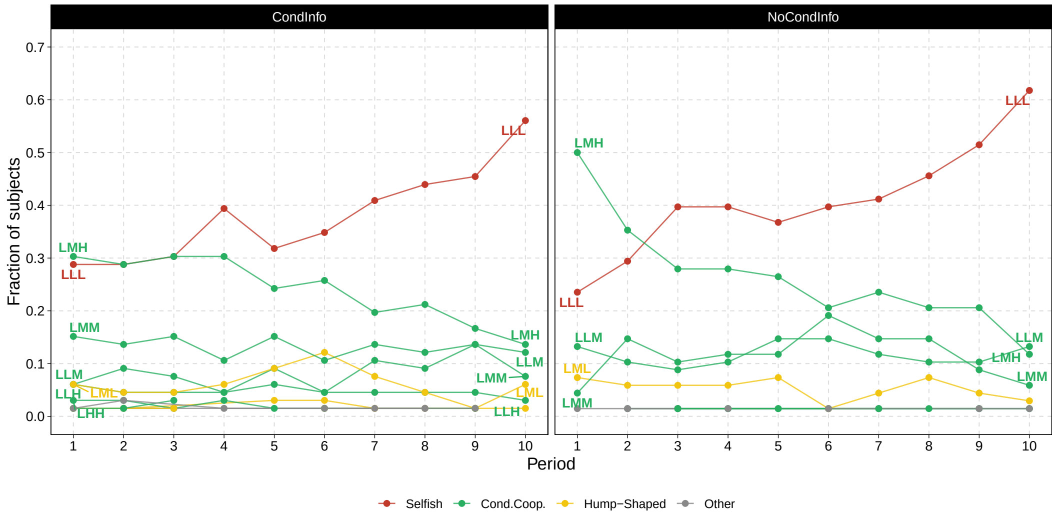

Figure 2.6: Conditional Strategies Over the Periods for Each Treatment

Result 3 - Perfect reciprocators have unstable preferences: As graphically illutrated, perfectly reciprocal strategy is less frequently used in the later periods. A multionial regression confirms this finding Table 2.3 while it suggest that the types other than perfect-conditional cooperators and selfish does not affect from game being played over time. Since subjects change preferences during the game, the notion that each individual has a "type" which is revealed at the beginning of the experiment loses most of its appeal. We used a simple metric to measure the stability of subjects according to their most frequent strategy: A subject's strategy in the first period is said to be stable if it is the most common strategy during the rest of the game. Figure 2.7 shows the average stability of the subjects according to this measure. The differences between stability levels of the types can be seen in Table 2.4.

	Model 1	Model 2
(Intercept)	0.63^{***}	0.34^{***}
	(0.05)	(0.05)
Period	-0.03^{***}	-0.02^{***}
	(0.00)	(0.00)
Treatment: *CondInfo*	0.00	-0.01
	(0.08)	(0.07)
Counterpart's Action (t-1):M	0.07^{***}	
	(0.01)	
Counterpart's Action (t-1):H	0.08^{***}	
	(0.01)	
Beliefs - Unconditional		0.16^{***}
		(0.02)
Beliefs - Response to L		-0.24^{***}
		(0.04)
Beliefs - Response to M		0.07
		(0.04)
Beliefs - Response to H		0.48^{***}
		(0.02)
AIC	8123.22	7148.93
BIC	8183.24	7223.96
Log Likelihood	-4053.61	-3564.46
Num. obs.	13390	13400
Num. groups: Subject	134	134
Num. groups: Session	6	6
Var: Subject (Intercept)	0.14	0.11
Var: Session (Intercept)	0.00	0.00
Var: Residual	0.10	0.09

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2.2: Mixed-Effects Logistic Regression: Dependent Variable: *isCondCoop*: The binary variable shows whether a subject is a conditional cooperator. Beliefs are aggregated: the expected transfer of the opponent and normalized to the interval of [0,1]

	Imp. Cond. Coop	Perf. Cond. Coop	Hump-Shp.	Other
(Intercept)	-2.47 (0.38)***	-3.48 (0.46)***	-4.62 (0.62)***	-7.40 (0.91)***
Period	-0.05 (0.03)	-0.12 (0.03)***	-0.08 (0.04)	-0.03 (0.08)
Treatment - *CondInfo*	0.17 (0.14)	-0.22 (0.17)	0.49 (0.24)*	-0.26 (0.50)
Beliefs - Unconditional	0.32 (0.40)	0.92 (0.45)*	-1.53 (0.71)*	1.51 (1.32)
Beliefs - Response to L	1.38 (0.71)	1.05 (0.80)	2.01 (1.01)*	9.95 (1.68)***
Beliefs - Response to M	2.29 (0.77)**	0.73 (0.89)	5.72 (1.15)***	0.31 (2.80)
Beliefs - Response to H	1.35 (0.52)**	5.23 (0.59)***	-0.09 (0.83)	-2.82 (1.89)
AIC	3049.29	3049.29	3049.29	3049.29
BIC	3194.90	3194.90	3194.90	3194.90
Log Likelihood	-1496.64	-1496.64	-1496.64	-1496.64
Deviance	2993.29	2993.29	2993.29	2993.29
Num. obs.	1340	1340	1340	1340

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2.3: Multinomial Logistic Regression. Dependent Variable: Conditional Type. Reference type: Selfish

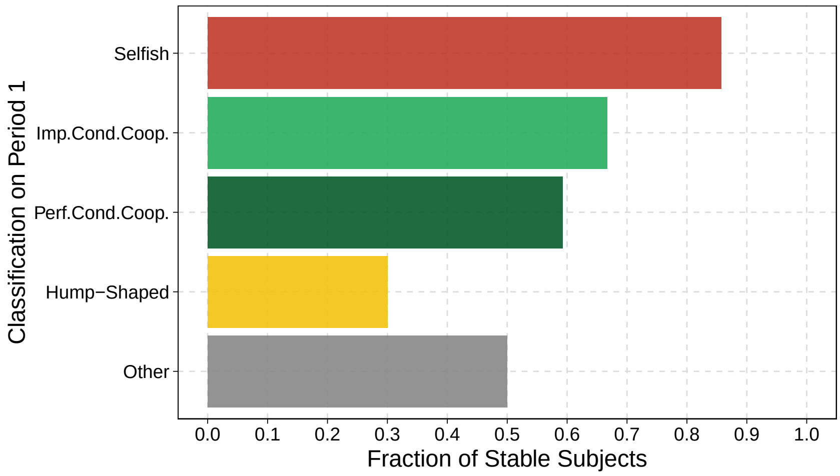

Figure 2.7: Stability of the Subjects' Decisions which are Classified According to Their Strategies in Period 1

2.4 Discussion

Our data reveal that when subjects' preferences are elicited after they had time to learn, reciprocal individuals tend to switch to selfish strategies, while the opposite transaction is rarely observed. This result suggest that, reciprocally altruistic motives play a smaller role than previously thought in the onset and subsequent decay of cooperation.

Dependent Variable: Stability	
(Intercept)	0.86^{***}
	(0.08)
Imp. Conditional Coop.	-0.19
	(0.11)
Perf. Conditional. Coop.	-0.26^{**}
	(0.10)
Hump-shaped	-0.56^{***}
	(0.16)
Other	-0.36
	(0.33)
AIC	178.87
BIC	196.26
Log Likelihood	-83.44
Deviance	27.26
Num. obs.	134

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2.4: Logistic Regression on the Stability of Types Based on the First Period. Baseline: selfish

A methodological implication of our study is about using conditional preferences as a sort of *index* for cooperative tendencies. Several studies has been using Fischbacher et al. (2001) method to elicit preferences for cooperation and treat them as stable types and then linking them with other results from several games or measures. Our results suggest that using this kind of approach might have some drawbacks. In our experiment, for instance, a fraction of the condititional cooperatiors switch immediately to selfish strategies after first few periods, while some of them stick with their choices for the rest of the interaction. If one-shot Public Good Game with strategy method is used to be a measure, then those types mistakenly would be assumed to have same tendency to cooperate which is clearly not the case.

Our result suggest that beliefs play a role together with learning in the appearent decay of cooperation. Obviously, the influence of beliefs to the conditional strategies is somehow unexpected. Individuals' conditional strategies are implemented only if a certain action is chosen by their counterpart, irrelevant of this action to be percieved as likely or not. Since we found no difference between our treatments in terms of strategies chosen by the individuals and the dynamics of the population, it is not certain for us, whether the beliefs and behavior has causal link between them. However, future research will reveal to what extent they are linked to each other.

Chapter 3

The Evolution of Conditional Cooperation

Abstract

Reciprocal altruism is considered to be a strong mechanism that can promote cooperation, if individuals are likely to interact. Experimental evidence shows that most of the people have reciprocal tendencies when they interact in social dilemma experiments, although the members of the population as a whole seem to be heterogeneous in terms of their conditional strategies. In addition to that, those who employ conditionally cooperative strategies also seem to have a certain degree of selfish bias; they tend to give less than what others give do. Cooperation by the individuals who employ such preferences is bound to collapse as the level of cooperation tends to decrease over time.

In this study, we investigate conditional types and their evolution in an iterated Prisoner's Dilemma , comparing different continuation probabilities, by using a computational model. In our setting, agents are characterized by their responses to each level of cooperation in a linearly extended Prisoner's Dilemma. By using repeated simulations, we estimate the likelihood of cooperation and the conditional strategies that are likely to succeed.

Our results show that, when the continuation probability is sufficiently large, full cooperation is achieved. In this case, the most successful strategies are the ones who employ an all-or-none type of conditional cooperation, followed by perfect conditional cooperators. In the intermediate levels of continuation probability, however, *hump-shaped* contributor types are the ones that are most likely to exist, followed by imperfect conditional cooperators. Those agents cooperate in a medium level of cooperation within themselves and each other.

Our results provide an explanation for the commonly observed hump-shaped strategy and imperfect conditional cooperators in experiments. Furthermore, a potential implication of our results is that the heterogeneity of conditional strategies might stem from the diverse interaction frequencies among real-world interactions.

Keywords: reciprocity | conditional cooperation | reactive strategies | hump-shaped contributors

3.1 Introduction

Humans are distinct in their cooperation capability, if not unique (See Bowles and Gintis, 2011; Wilson, 2012). As a species, we heavily depend on the mutual cooperation of individuals. At first glance, from an evolutionary perspective, it seems counter-intuitive for individuals to care for the others: if everybody is cooperating in a population, any fraction of selfish individuals would do better than those who care, and would therefore have a higher evolutionary fitness. As it would be easy to exploit the cooperators in such a population, it is just matter of time for cooperation to be eradicated. Then, how come cooperation is so common in some species including us humans?

Several explanations have been put forward to explain this puzzle.[1] One of the most prominent answers has been reciprocal altruism. The earliest systematic thoughts on the evolution of reciprocal altruism can be found in the writings of the Russian thinker Kropotkin in the early 20th century. In his book *Mutual Aid*, he discusses the evolutionary success of solidarity with some examples from animal groups and human societies in historical perspective (Kropotkin, 1902). "*(...) Natural selection continually must eliminate (anti-social instincts), because in the long run the practice of solidarity proves much more advantageous to the species than the development of individuals endowed with predatory inclinations*", he asserts (Kropotkin, 1902, p. 18). His arguments today can be interpreted as a mixture of different mechanisms such as reciprocal altruism, group selection and kin selection. It was not until more than half a century later that the biologist Robert Trivers coined the term *reciprocal altruism* and explained its role in the context of cooperation by way of a formal model (Trivers, 1971). By using Hamilton's methodology of kin selection (Hamilton, 1964), Trivers showed that, given that there is a

[1] See Nowak (2006) for a summary of some of these mechanisms.

positive probability for individuals to interact again, if the benefit of an aid is sufficiently greater than its cost, then cooperation of unrelated individuals can arise. Trivers's treatise on reciprocity has been highly influential, especially in the field of theoretical biology.

In game theory literature, Folk theorems show that, if the threat of defection is present in the future, cooperating today is more beneficial; therefore, a cooperative strategy in the stage game, as well as any other individually rational strategy, can be the equilibrium of the supergame (Friedman, 1971; Fudenberg and Maskin, 1986). In addition to that, Axelrod's simulations on the Prisoner's Dilemma support those theoretical results by showing that the reciprocal *tit-for-tat* is a strong strategy that allows cooperation while it is immune to exploitation by selfish players (Axelrod, 1980a; Axelrod, 1980b; Axelrod and Hamilton, 1981).

With these scientific accounts, a rather optimistic picture of cooperation through reciprocity has been drawn. The main dynamics of cooperation through reciprocity can be summarized as follows: reciprocators cooperate among each other as they reciprocate kindness over time. Therefore they receive a higher payoff than the selfish types who meet other selfish types. And when a reciprocator meets a selfish type, the reciprocator responds selfishly after the first defection. Thus, the exploitation of a reciprocator by a selfish individual is not sustainable when the probability of continuation is sufficiently high.

Nevertheless, with the rise of experimental economics, empirical evidence on the Prionser's Dilemma and the Public Goods Game draws a rather different picture: the contributions tend to decline over time, whether subjects play the game within the same group or matched with others (Selten and Stoecker, 1986; Andreoni, Miller, et al., 1993; Cooper et al., 1996; Ledyard, 1994; Kim and Walker, 1984; Isaac et al., 1985; Andreoni and Croson, 2008). This declining trend in the Public Goods Game is often, and arguably best, explained by conditional preferences. Clear evidence is provided by Fischbacher et al. (2001), who collects conditional choices of subjects to the common pool by using the strategy method. According to their results, most of the participants reveal a preference pattern, in which they will only contribute if the others contribute as well, albeit with a *selfish bias*; they tend to contribute less to the public pool than the others. Those type of conditional cooperation is not sufficient to sustain cooperation, as contributions tend to decline over time(Fischbacher and Gächter, 2010).

In this study, we aim to investigate the conditional preferences in the framework of Fischbacher et al. (2001), using computational methods. We examine the evolutionary success of those conditional strategies in a social dilemma with a simplified framework. We create populations of agents with the types drawn from of the set of all conditional strategies within our framework. Each agent randomly pairs with another agent to interact in this social dilemma, and continues to interact with probability δ. We let an evolutionary reproduction mechanism based on the success of the strategies shape the population structure and we examined success of those strategies for different continuation probabilities δ.

Our results can be summarized as follows: selfish and close-to-selfish conditional cooperators are likely to be successful when the repetition probability is relatively low. Depending on the value of continuation probability, imperfect conditional cooperators who only fully cooperate when the opponent fully cooperates, perfect-conditional cooperators, and agents with hump-shaped contribution schedule are relatively successful when the repetition probability is sufficiently high. Unconditionally cooperating strategies are unlikely to survive even at high probabilities of repetition. And finally we conclude that the initial move aligned with the conditional strategy plays a crucial role in the survival of those conditional strategies.

3.2 Literature Review

Conditional cooperation has been attracting interest in economics, especially after the seminal work of Fischbacher et al. (2001) who conducted a Public Goods experiment that gave subjects the chance to choose the contribution amount to the public pool according to the average contribution of the others. According to the contribution schedule, they classified subjects as follows: *conditional cooperators*, who increase their contributions with the increasing amount of average contribution of others; *selfish players*, who contribute nothing to the public pool regardless of what others do; and *hump-shaped contributors*, who increase their contribution up to a certain level, most commonly to halfway to the maximum contribution, only to decrease their contributions after that level, as the contributions of others continue to increase. *Unconditional cooperators* were virtually absent in their experiment. Fischbacher, Gächter and Fehr concluded that most of the subjects are selfish-biased conditional cooperators. In a follow-up study, Fischbacher and Gächter (2010) showed the dynamics of such contribution patterns, which predict the decline in Public Goods Games. Those conditional strategies defined by Fischbacher can be investigated in the class of reactive strategies, as they only consider the previous move (or, in the simultaneous case, the current move) of the player(s) with whom they interact.

Possibly because rational and self-interested agents have been central in economic methodology, early models based on conditional strategies were investigated in disciplines other than economics, which employ game-theoretical approaches, such as biology and political science. Axelrod's research showed the success of *tit-for-tat*, a strategy beginning with cooperation in the initial interaction and then copying its opponent's from the previous period (Axelrod, 1980a; Axelrod, 1980b; Axelrod and Hamilton, 1981). A tit-for-tat strategy in a Fischbacher-Gächter framework would be equivalent to a perfect-conditional-cooperation strategy that matches the individual's contribution to the contribution by others. One should expect this kind strategy to be employed in real-life decisions due to the success of tit-for-tat strategy. In the experiments, however, perfect-conditional-cooperators are outnumbered by imperfect conditional cooperators who tend to give less than what others give. Reciprocity by imperfect conditional cooperators not enough to sustain cooperation, not just within other commonly observed conditional types, but also when they interact among each other (Fischbacher and Gächter, 2010). Moreover, later results showed that the tit-for-tat strategy was vulnerable when mistakes happen: when an accidental decline occurs, tit-for-tat was not able to reestablish cooperation back again, unless another mistake corrected it (Hirshleifer and Coll, 1988; Selten and Hammerstein, 1984; Fundenberg and Maskin, 1990; Nowak and Sigmund, 1993).

A related line of research with conditional strategies was mostly investigated in biology. So-called "variable investment models" provided more insights into successful conditional-like strategies in social dilemmas. These models differ considerably from conditional strategies defined in a Fischbacher-Gäcther framework, as well as from each other by their assumptions; therefore it is not straightforward to compare their results.[2] For instance, in Doebeli and Knowlton (1998), the agents are characterized by two parameters; a parameter that defines the unconditional altruism, and another defining the amount of reciprocation proportional to the difference between one's own altruism and that of the other. In their setup, they use a continuous altruistic investment game in which the marginal benefit of

[2]See Sherratt and Roberts (2002) for a review of some early papers.

altruism decreases. Their evaluation of the success of their strategies takes pairwise comparison into account with the mutants of that strategy. They show that, unless a spatial structure is assumed, selfish types are the most successful ones. Roberts and Sherratt (1998) use a different agent definition in which costs and benefits are linear and investments can increase. They take representative strategies for a tournament: *Non-altruistic* agents, who do not invest at all; *Give-as-good-as-you-get* strategies, which invest as much as the other; *raise-the-stake* strategies, amplifying the investment, *short-changer* strategy, which gives less to the other than what the other gives, *all-or-nothing* strategy, which gives an amount and amplifies it only if other reciprocates as much or more. Roberts and Sheratt find that the raise-the-stake strategy is a stable cooperator as it cautiously invests but repeated interactions allow it to reach high cooperation levels.

Wahl and Nowak (1999a) and Wahl and Nowak (1999b) provide a framework of reactive strategies under adaptive dynamics by providing a linear and continuous extension of the Prisoner's Dilemma in which different amount of contributions are possible. In their framework, they define a reactive strategy with three parameters: the initial move, the slope of the reaction function, and the intercept of the reaction function. Their results suggest that the initial move is a decisive factor in terms of success of a strategy. Moreover, Wahl and Nowak conclude that strategies that are generous, uncompromising and optimistic (in terms of initial move) are the most successful ones. An important result of those studies is that they demonstrate cycles of cooperation and defection over time. Indirect invasions deter reciprocal strategies which are normally resistant to direct invasion: more cooperative and/or compromising strategies are able to invade reciprocal strategies as those types are exploitable. That allows the invasion of the population by selfish types.

van Veelen et al. (2012) uses a generalized computational model. They run simulations of a Prionser's Dilemma game with finite automata. Their strategies can evolve to be reactive strategies or strategies that consider several rounds of histories depending on the complexity of the strategy generated by mutations. Moreover, their model investigates different levels of assortativity in the population structure. They show that reciprocity and population structure can jointly work in favor of reciprocal strategies. However, reciprocity itself is not sufficient to create and to foster cooperation with random interactions.

Considering these theoretical and computational results on conditional cooperation and reactive strategies, the following question arises; if we consider the set of strategies with those conditional strategy classifications defined by Fischbacher et al. (2001) framework, how likely would it be for cooperation to arise, and which particular strategies would be more successful than the others?

The two closest approaches to our study are Szolnoki and Perc (2012) and Zhang and Perc (2016). Szolnoki and Perc (2012) use a spatial structure to investigate conditional cooperation in Public Good Games. Their agents condition on the number of cooperators rather than the strategies themselves. The authors found that the more cautious conditional cooperators enable high levels of cooperation and limit the outreach of defectors by surrounding them spatially. And a recent paper by Zhang and Perc (2016) investigates the evolution of conditional types in a Public Goods Game in an agent-based model. The strategies their agents use are a combination of continuous and discrete functions. In authors' setup, a multilevel selection procedure is used. They found that the most successful type is the conditional cooperator type that contributes nothing up to a mid-level contribution by others, but increases its contribution gradually to the full cooperation when others agents' average contribution is maximum. This type of strategy can be considered as “all-or-nothing” type of strategy.

As we draw the general picture, there are certain regularities and mixed results in the literature. Those related with our research questions can be summarized as the following:

- Cooperation and defection strategies are likely to oscillate in most settings, especially without certain assumptions in the population structure. In Prisoner’s Dilemma game, no pure strategy is evolutionary stable, and reactive strategies are vulnerable to indirect invasions.[citation here] Therefore, those strategies that use a mixture of actions need further investigation.
- When there is room for mistakes, it is often unclear how the results obtained by simulations align with theoretical expectations.
- There is a certain degree of mismatching between the actual behavior in experiments and the theoretical predictions in the framework of conditional strategies.

Therefore our study aims to fill this gap by demonstrating the success and estimating the likelihood of conditional strategies.

3.3 Methods

Our model has certain differences compared to those models in the literature which we referred in the previous section. First, we choose a setup that is closer to a standard economics framework, in the sense that we do not assume a population structure, but instead use uniform random matching procedure for our agents. However our simulations can easily be extendable to such population structures. Second, we use a minimal setup which allows us to observe conditional strategies, ad defined by Fischbacher & Gächter framework, and which avoids unnecessary computational complexities. We use an iterated Prisoners Dilemma with three strategies and discrete linear response functions that do not assume or limit the shape of our response functions. Moreover, we also control for the initial reactions, as it is often stressed that they play an important role on the path of the dynamics. In this section, we start by explaining the social dilemma framework we use, and then we give details of the evolutionary model we use. And we finally describe the outcome variables we are interested in.

3.3.1 Linear Extension of the Iterated Prisoner's Dilemma

We use an extension of the Prionser's Dilemma game. In this extended game, there are three actions, representing different cooperation levels: L(Low) refers to no cooperation/defection; M(Medium) refers to an intermediate cooperation level; and H(High) refers to full cooperation.

The cost of cooperation is c, while the benefit of cooperation to the opponent is b, with the constraint $b > c > 0$. If an agent plays the strategy L, he keeps his endowment for himself and no change will occur in payoff of the opponent. If agent plays the strategy H, he pays the cost c and his opponent gets the benefit b. If the player plays M, he will only pay the half the amount of the cost of full cooperation but the benefit to the opponent will be half as much as well, i.e., the agent pays a cost of $c/2$, while the opponent gets the benefit $b/2$.

If the agents were to decide simultaneously, the game matrix would be as shown in Table 3.1. For our simulations, we used the following parameters: the cost of cooperation c is equal to 1, the benefit of cooperation b is equal to

2, and we set the base payoff to 1 to avoid negative payoffs. The normal-form representation of the game with those parameters is shown in Table 3.2.

		Player 2 L	M	H
Player 1	L	$(0,0)$	$(\frac{b}{2},-\frac{c}{2})$	$(b,-c)$
	M	$(-\frac{c}{2},\frac{b}{2})$	$(\frac{b-c}{2},\frac{b-c}{2})$	$(b-\frac{c}{2},\frac{b}{2}-c)$
	H	$(-c,b)$	$(\frac{b}{2}-c,b-\frac{c}{2})$	$(b-c,\, b-c)$

Table 3.1: Extended Prisoner's Dilemma Game

		Player 2 L	M	H
Player 1	L	$(1,1)$	$(2,0.5)$	$(3,0)$
	M	$(0.5,2)$	$(1.5,1.5)$	$(2.5,1)$
	H	$(0,3)$	$(1,2.5)$	$(2,2)$

Table 3.2: Extended Prionser's Dilemma Game with Simulation Parameters b=2, $c = 1$ and the Base Payoff 1

In order to investigate conditional strategies in terms of the effect reciprocity on cooperation, we used a sequential game setting based on the iterated Prisoner's Dilemma. In our game, for each pair, one of the players is randomly selected as the first mover and makes the initial move. At this point, the agent will bear the cost of his or her action, and the opponent will benefit from this action, if any.

We use an infitely repeated game approach to the iterated Prisoner's Dilemma Game: After each round, the game continues with the probability of repetition δ.

In the first round, the first movers start with their initial move. Then if the game continues, second movers respond to what their match did in the previous round according to their conditional strategies. If the game further continues, the first-mover agents in the first round similarly responds to that action with their own conditional strategy.

3.3.2 Conditional Strategies

In our model, an agent is characterized by his or her conditional strategy. As our extended Prisoner's Dilemma has three possible actions (L, M and H), a conditional strategy has three components in addition to the initial strategy. We represent each strategy with four letters; the first letter separated from the rest with a hyphen. The first letter corresponds to the initial move of the strategy, where the other three letters denote the conditional responses to L, M, H respectively. Since, in our context, there are three possible actions for four different situations, in total there are 81 possible strategies(3^4). The set of all possible strategies and their classification can be found in Chapter B.

Figure 3.1 is a demonstration of two conditional strategies in interaction. The left side shows the actions of the first agent represented by the blue color, while the right side shows the actions of the agent represented by the red color. The blue circle refers to the initial move of the left agent, while the blue arrows, pointing from the actions of the other agent to the player's own actions, represents a conditional reaction. Similarly, for the red agent, the red circle represents the initial movement, and the arrows represent the conditional responses.

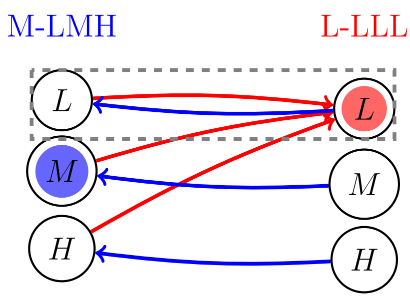

Figure 3.1: A Demonstration of the Interaction of Two Agents with Respective Conditional Strategies $M - LMH$ and $L - LLL$

As can be seen from the figure, the blue player has the action M as the initial move, and respectively he/she plays L if the opponent plays L, M if the opponent plays M, and H if the opponent plays H. Therefore, this type of the agent is denoted as $M - LMH$.

To give an example of the interaction, if the blue player is selected as the first-mover, it starts with the action M. The red arrow from that node points to action L of the red player, therefore the red player responds with L. If the game continues,the blue player responds with L and if it continues further, the red player responds with L, and so on.

3.3.3 Computational Model

In terms of the procedures of population generation, reproduction and mutation, we use a similar methodology with van Veelen et al. (2012). First, we generate a population with a fixed size of 200 agents. The type of the agents in this population are sampled from all 81 possible types. Each agent lives for one generation. In each generation, agents are matched in pairs randomly. After they are paired, the agents play the extended iterated Prionser's Dilemma we described in the previous sections. The interaction is repeated with probability δ. To reduce computational complexity, we sampled the number of interactions from a geometric distribution for each matching in each generation.[3] For each interaction, we normalize the payoffs by dividing the total payoff by the number of interactions in order to fix the effects of different delta values. After the interaction by two individuals are stopped, we resample the population according to payoffs; each agent is expected the have the offspring in the next generation with the probability that is equal to the proportion to its payoff to the total payoff. Therefore, agent $i \in N$ has the probability of resampling in the next generation $\dot{p}$:

[3]The expected number of interactions for a given δ is $E[T] = \frac{\delta}{1-\delta}$. As the first interaction occurs with certainity in our setting, the expected number of interactions is $1 + \frac{\delta}{1-\delta} = \frac{1}{1-\delta}$.

$$\dot{p}_i = (\pi_i / \sum_{j \in N} \pi_j)(1 - p_M),$$

where π_l denotes the normalized payoff of individual l, N denotes the set of agents, and p_M denotes the probability of mutation. Each type has the same probability of taking place in the next generation through mutations.

We use Monte-Carlo method for our investigation: We repeat each simulation with the specific parameter for 500 times independently. That allows us to obtain the mean frequency of each action and each strategy in each parameter and generation over the total number of simulations. Then, these frequencies can be interpreted as the probability of an action/strategy to exist in a given parameter.

Stage	Details
Population Generation	Number of Agents: 200 Uniformly from all possible types
Matching	Two players: random matching
Interaction	Extended iterated Prisoners Dilemma Repetition with probability δ
Reproduction and Mutation	Resampling proportional to normalized payoff Uniform random mutation with a fixed probability Ran for 5000 generations
Resampling	Regeneration of the population with the same parameters for 500 times

Table 3.3: Summary of the Computational Stages

In our model, each agent has a probablity of making a mistake $p = 0.005$. This means, either as the first mover in the first interaction, in the later interactions, an agent plays a random strategy. Our results show that mistakes we introduce has minor effects on the rate of cooperation. The comparision of the simulation results between the case with mistakes and without them can be seen in Figure B.1.

As the repetition probability δ is the key parameter we are interested in, we covered a range of values from 0.5 to 0.95 with increments of 0.05. The results we have obtained are reported in the next section.

3.4 Results

To interpret the results we have obtained, first we start by demonstrating three single instances of our simulations for different delta values. Figure 3.2 shows the fraction of the actions that are being played in each generation, while Figure 3.3 shows the distribution of types in the same interaction.

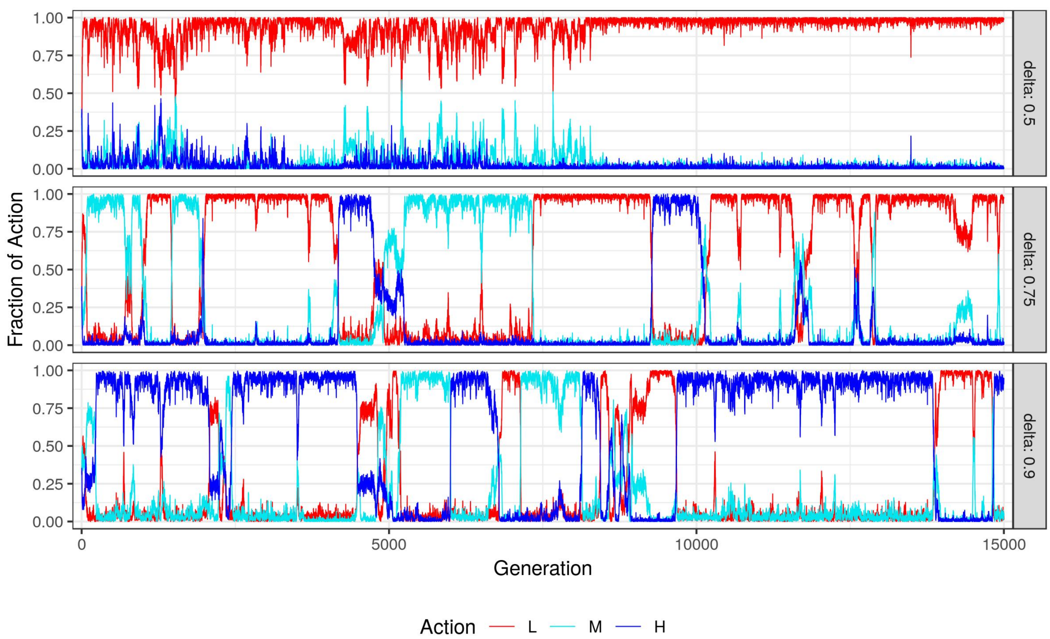

Figure 3.2: Fraction of Actions During Three Instances of Simulation with Differnt Continuation Probabilities

As can be seen from Figure 3.3, no type dominates the population for a long time. However, when δ is low, defection(L) outperforms cooperative strategies. In this case the types that are taking over the population are neutral mutants of the selfish type. When δ increases, occasional oscillations occur both in types and also in the actions. A single type takes over the population and stays as the most common type until another type with direct or indirect mutation takes over. This result is not unexpected, as no type is resistant to indirect invasion in a Prionser's Dilemma (van Veelen et al., 2012; García and van Veelen, 2016; García and van Veelen, 2018). But it gives us the intuition on how to interpret our result of resampled simulations. As cooperation and defection can occasionally occur, we should interpret our result in a probabilistic way rather than an expected state of a population.

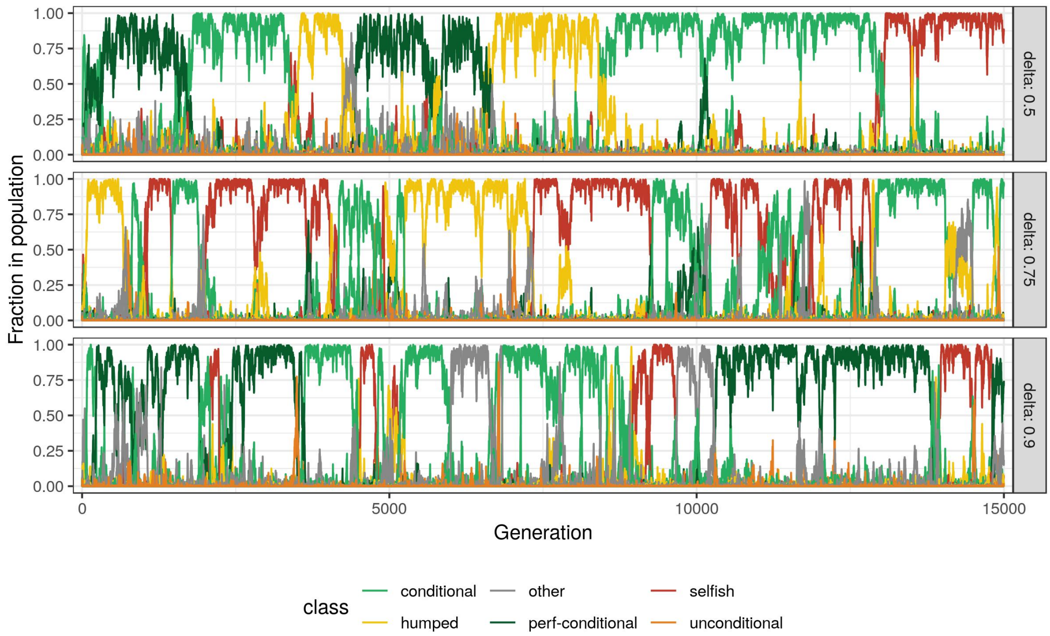

Figure 3.3: Fraction of Strategies During Three Instances of Simulation with Different Continuation Probabilities

Result 1 - Full cooperation is more likely to be obtained only at high values of continuation probability: Figure 3.4 shows the average fraction of each action; and similarly, Figure 3.5 shows the average fraction of each type for different values of repetition probability δ. Each value shown is an average of 500 simulations that ran over 5000 generations. Moreover, to reduce the effect of occasional drifts, we averaged last 2000 generations for which our simulations showed no great variation on the average values.

As δ increases, the existence probability of M and H gradually increases and the values of δ in which cooperation is more likely than defection is obtained only when delta is above a certain threshold. (For our parameters, $\delta > 0.75$, which is well above the theoretical threshold c/b. Hence, reciprocity is unlikely to promote cooperation below those values. Moreover, mid-cooperation, although being Pareto-inferior, is more likely to be obtained for those values.)

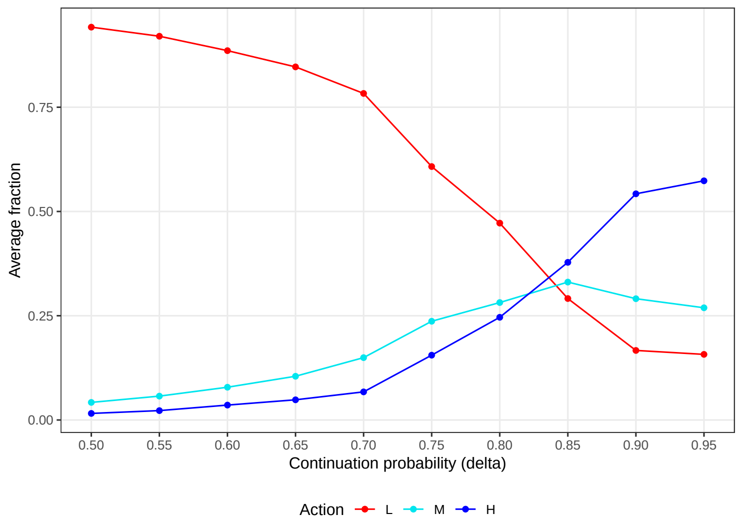

Figure 3.4: Average Fraction of Actions Over 500 Repetitions for Different Values of Continuation Probability δ

Focusing on Figure 3.5, we see that conditional types that are closer to selfish types are likely to exist in relatively lower values of δ, though it is unlikely for them to foster cooperation, but they rather likely to survive because they are neutral mutants when defection is a the common action in the population.

Result 2 - All-or-nothing cooperators are more likely to exist than perfect conditional cooperators: Figure 3.5 shows that, in all continuation probabilities, all-or-nothing cooperators with an optimistic start($H - LLH$) are more likely to exist than perfect conditional cooperators

$(H - LMH)$. Those two types are equivalent if we remove the medium level of cooperation. In this case, they both play a tit-for-tat strategy.

These strategies would do equally well when they play against each other in a deterministic setting: They would engage in full-cooperation(H), which would be adventageous for both types when the conditional cooperation probability is sufficiently high. However, when mistakes happen, all-or-nothing strategy is likely to exploit mistaken mid level cooperation(M) by others.

Result 3 - Initial move of a strategy is a determinant of the success of a conditional decision: Consider three strategies which are identical in their responses to the counterpart, but differ in their initial move: $L - LLH$, $M - LLH$ and $H - LLH$. All of these types, in an experimental study, would be classified as conditional cooperators. Our evidence shows that their success is highly dependent on the initial move. For instance, $L - LLH$ is relatively successful when δ is low and cooperation is unlikely. In this case, an opponent that plays H is unlikely to exist in the population, and the reaction of this strategy is L for all other strategies. This strategy does as good as the selfish the selfish strategy. However, agents using this strategy often fail to cooperate within each other, unless a mistake leads to high cooperation, as none of the two interacting parts initiate cooperation. The twin strategy that is successful in cooperating is $H - LMH$, but when the repetition probability is low, the first move of this strategy would cost a significant fitness that is impossible to tolerate with the cooperation within itself. When the δ is low and the population lacks a structure that gives a higher probability for similar agents to interact, this strategy fails to survive. The strategy $M - LLH$ has strong disadvantages both where δ is low, and it is sufficiently high to sustain cooperation.

Result 4 - Hump-shaped strategies are relatively successful: An interesting result we obtained is the relative success of hump-shaped cooperators. Such types of strategy are constantly observed in empirical studies, while the arguments regarding this type of behavior in those studies are still have been far from convincing.

In our simulations, we have two types of humped-shaped strategies that are likely to exist: $L - LML$ and $M - LML$. The former one is relatively successful due to its proximity to selfish strategy and virtually behaving as the same as the selfish agents when no cooperation is common in the population.

The latter one, however is a more interesting case. The reason begind the success of $M - LML$ strategy is due to its ablility to coordinate in medium level of cooperation with a smaller cost than the other strategies. Moreover this strategy can exploit high-contributors as well. When the continuation probability δ is sufficient for cooperation to occur but still risky for high cooperation, this hump-shaped strategy is relatively successfull.

Figure 3.5: Average Fraction of Strategies Over 500 Repetitions for Different Values of Continuation Probability δ. The Figure Includes the Strategies Which Consists at Least 10% the Population for at Least One Delta Value.

3.5 Discussion

In this study, we investigated the emergence of cooperation and relative success of conditional strategies in an extension of iterated Prionser's Dilemma. Some of the results we obtain provide supporting evidence for the previous results in a different setting. For instance, the oscillations between defection and two levels of cooperation were expected in the light of previous studies (Wahl and Nowak, 1999b; Nowak and Sigmund, 1989; Bendor and Swistak, 1995; Imhof et al., 2005; van Veelen et al., 2012; García and van Veelen, 2016; van Veelen et al., 2012). Also, the initial strategy is indeed decisive for the success of a type (Wahl and Nowak, 1999a; Wahl and Nowak, 1999b). Our results suggest that the initial moves should be in accordance with the action where a strategy is most successful within itself.

The most successful types considering all possible continuation probabilities are: selfish and relatively selfish types that start with defection(L); hump-shaped and conditional cooperator type which has the conditional response LMM that start with M; and all-or-nothing and perfect-conditional-cooperator types that start with H. Those types outperform other types who have the same conditional strategy but a different initial move. Those results suggest that strategies with a non-aligned initial move fail to reach the interaction cycle where they would profit most.

A particularly surprising result concerns hump-shaped contributors. Those types are often observed in experiments in which conditional preferences are elicited. They are indeed the most common types after conditional cooperators and selfish types in those experiments. It is hard to rationalize this kind of strategy which can be counter-intuitive. At a first glance as they would fail fully to exploit others while not being able to cooperate at a maximum level, but rather stuck in a medium level of cooperation. However our results suggest that as they are able to exploit mistakes made by their opponents better than the cooperative types and they do relatively better than non-cooperative types, it gives a relative advantage where the continuation probability is moderate. If we also consider the case where repetition probability changes over time, we can expect those types to do relatively better than the others. Our setup might be misleading though; since we have three levels, a monotonic increase or a monotonic decrease is not possible around the intermediate cooperation. What should one expect if there were more levels of cooperation? Possibly the answer depends on the structure of mistakes: If we assume mistaken moves to similar actions are more likely, then mistaken moves to actions that are far from the outcome of the response function,and then we might expect those monotonic increments and decreases. Due to computational complexities of this expansion creat, we seek for further research to confirm this hypothesis.

Finally, our result might provide some explanation about the heterogeneity of conditional strategies we observe in experiments. In our setting, different type of strategies arise and different continuation probabilities: selfish types when δ is small, perfect-conditional-cooperators and all-or-none type of conditional cooperators when δ is high and hump-shaped and imperfect conditional cooperators when delta is in-between. Though it is possibly a strong claim that those conditional preferences are direct consequences of evolution, their success in different continuation probabilities are evident. Therefore mixing individuals with different histories, either at a cultural sense or at an evolutionary sense, might result in such heterogeneity.

Chapter 4

Presumptive Reciprocity in Dictator Games

with Luciano Andreozzi and Marco Faillo

Abstract

The Dictator Game was initially introduced as a test for pure, non-reciprocal altruism. The large experimental evidence in the early literature revealed that subjects often share a fraction of their endowment; in addition, a considerable number of them even choose an equal split. In the years that followed, much criticism was levelled against these results. Some claimed that the evidence collected in the lab has little external validity. The kind of generosity we observe in the Dictator Game does not seem to be compatible with the giving patterns we observe in real life. Outside of the lab, donations to strangers usually target carefully chosen individuals or groups in need of help. Others observed that giving in the Dictator Game can be easily manipulated with trivial changes in the framework in which the game is presented. This originated a theoretical literature that tries to explain dictator giving in terms of other motives besides altruism, such as self-image and reputation.

In this essay, we take a more traditional stance. We argue that part of giving in the dictator game may in fact be explained in terms of altruism. However, it is a type of reciprocal altruism which is based on the presumption that the recipient would have behaved altruistically as well, if the roles were reversed. To do so, we use the strategy method to elicit subjects' preferences in a Dictator Game with randomly assigned roles. We ask subjects to choose a level of giving, conditional on possible levels of giving of the other player. We compare our results from two treatments: (i) when the recipient is a peer in the laboratory; (ii) when the recipient is a member of a low socioeconomic group. The presumptive reciprocity hypothesis predicts that when subjects play against their peers they should reveal conditionally altruistic preferences. Instead, they should be more likely to be unconditionally altruistic when playing against a subject in need. We find that these intuitive predictions are only partially borne out by the data. Whether giving is directed to a person with similar socioeconomic status, or it is directed to a person with low socioeconomic status, most subjects reveal conditionally altruistic preferences. Unconditional altruism seems to be rare for both treatments. However unconditional altruists thansfer a significantly higher amount to the members of the group with a low socioeconomic status.

Our results suggest that a large part of the altruistic behavior observed in dictator game can be explained by presumptive reciprocal altruism, while unconditional altruism has a relatively limited role even in those cases in which it should.

Keywords: altruism | dictator games | reciprocity | presumptive reciprocity | social preferences | socioeconomic status

4.1 Introduction

The Dictator Game has for a long time been considered the sharpest tool to isolate pure altruism from reciprocity (Forsythe et al., 1994). In the simplest terms, the Dictator game is a choice problem in which a participant, *the dictator*, allocates an amount of money between himself and another person, *the recipient*. The recipient, in a passive role, receives the amount that the dictator allocates him. As it is frequently argued, since only one player is active, the game can be used to test the players' pure preferences over possible outcome distributions, without the inference of other motives, such as reciprocity.

Up to now, a large number of Dictator Game experiments have been conducted.[1] In most replications of the experiment, subjects share around 30% of the endowment with the recipients on average. More than 15% of the subjects choose an equal split, and more than 60% of subjects give a positive amount to the recipients. There are a number of reasons to find these results perplexing. First, this reveals a degree of altruism (or inequity aversion) that seems hard to reconcile with observed facts, such as actual donations to charities (Benz and Meier, 2008). Second, giving money to randomly chosen strangers seems to be absent among the repertoire of human behavior, however altruistic. Individuals usually direct their donation carefully to people in need. Third, evidence shows that apparently innocuous manipulations on the experimental framework change the subjects' decision dramatically in ways that are incompatible with rational choice (Hoffman et al., 1996; Dana et al., 2006; Dana et al., 2007; List, 2007; Bardsley, 2008; Lazear et al., 2012).

It is argued that these puzzling results in the Dictator Game can be explained by the social norms and self-image concerns (Andreoni and

[1] See Engel (2011) for a meta-study that covers the studies between 1992 and 2011.

Bernheim, 2007; Krupka and Weber, 2009; Tonin and Vlassopoulos, 2013). In a standard Dictator Game the perceived social appropriateness of each allocation affects their likelihood of being chosen. Trivial manipulations of the game change the behavior as such manipulations change the perception of each option. In this paper we pursue a different line. We argue that positive giving in the Dictator Game can be best understood as a form of reciprocal altruism, similar to the one usually observed in other games such as the Gift Exchange Game, the Trust Game and the Public Goods Game.

This possibility is usually overlooked, as apparently in the Dictator Game the recipient is idle, which excludes the possibility of a dictator's decision to give being motivated by reciprocity. Even models that assume reciprocally altruistic preferences (such as Falk and Fischbacher (2006) and Charness and Rabin (2002)) are usually complemented with a term that represents a "pure" concern for the fairness of the outcome.

A vast amount of empirical evidence has been showing the importance of reciprocity in experiments by using several economic games (Berg et al., 1995; Fehr et al., 1993; Charness, 2004; Clark and Sefton, 2001b; Gächter and Falk, 2002). For instance, the Trust Game (Berg et al., 1995) and the Gift-Exchange Game (Fehr et al., 1993) show the strong reciprocal tendencies in the lab. However, few studies are interested in the role of any kind of reciprocal considerations in the Dictator Game.For instance, Diekmann (2004) uses a sequential Dictator Game in which the recipient in the first round is the dictator in the second round. They compared three predetermined giving amount by dictators in the first round. Their results show that there is a strong correlation between the given amount of the first round and the modal giving in the second round.

Ben-Ner et al. (2004) collect the decisions of dictators, right after they were recipients in the previous round. They compare the case when the subject is playing with a partner, i.e., when the recipient is the dictator of the previous round, to the case when they are matched with another person. The giving in the second round is highly correlated with the amount they received in the partner matching, while in the stranger the matching, this correlation is insignificant.

Servátka (2009) focuses on indirect reciprocity. He investigates the effect of reputation in the Dictator Game, comparing two treatments: with and without the information about a previous round. The results show that the transfers are significantly higher to recipients who have a positive reputation based on their generosity, compared to recipients about whose past behavior is not known.

Considering these results, we expect a natural way of reciprocal reasoning possibly to play a role in the Dictator Game. A subject's decision to donate in the dictator's position might stem from his implicit belief that the recipient would donate as well if the roles were reversed. We refer to this kind of reasoning as "*presumptive reciprocity*", a reciprocal act based on the expected behavior of the person in the decision-maker's position (Çelen et al., 2017).

Perhaps the most direct way of experimentally checking this property of subjects' preferences is the one pioneered by Fischbacher et al. (2001) to investigate reciprocal preferences in a Public Goods Game. It amounts to simply asking each subject's contribution, conditional on the other subjects' average contribution. Subjects who are conditionally cooperative display contributions that are an increasing function of the other subjects' average contribution. We use the same experimental methodology to check the hypothesis that a dictator's willingness to give is a function of his beliefs about what the recipient would have done if he happened to be the dictator. Purely altruistic (Andreoni and Miller, 2002; Fisman et al., 2007) or inequity-averse (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) subjects would display a constant giving pattern; however, reciprocally motivated subjects will show an increasing contribution. For instance, Levine (1998) provides a formulation of reciprocal altruism incomplete information. The main ingredient of the model is the idea that individuals vary in their degree of altruism, and want to be more altruistic towards other altruistic individuals. Given a two players game in which payoffs are m and y, the decision maker maximizes

$$V(m, y) = m + (\theta_m + \lambda\, \theta_y)y, \tag{4.1}$$

where $\theta_i \in [0, 1)$ is the coefficient that determines how much player $i = M, Y$ cares about the payoff of the other player.[2] $\lambda \in [0, 1)$ is a parameter that determines how much a subject's altruism is influenced by the other player's altruism. When $\lambda = 0$ a player's altruism is unconditional. For $\lambda > 0$ a player is more altruistic towards altruistic players. The crucial assumption is that θ_i is private information and only the strategy chosen by an individual can be observed. In games with observable moves like the ultimatum game, reciprocal altruism gives rise to signalling, as players will choose strategies that influence other player's beliefs about their own type. This is not the case, however, for simultaneous moves games like the PGG and the DG.Because preferences are assumed to be linear in money, this model cannot explain partial transfers. In principle this could be fixed by letting utility be a non linear function of money.

Applied to our setting, if subjects have these type of reciprocal preferences, they should increase their contribution according to the type of the opponent signalled by the transfer in the Dictator game. In a not mutally exclusive way, if there is room for purely altruistic preferences in the Dictator game, any

[2] The original model contains a normalizing factor that we omit.

disparity in the background income should be noted. To test this hypothesis, we run a treatment in which recipients have a lower socioeconomic status. In line with the literature (Eckel and Grossman, 1996; Engel, 2011; Benz and Meier, 2008), we should expect that the information about the lower socioeconomic level of the recipient to increase giving in the Dictator Game. By collecting choices in strategy method, though, we can observe how conditional giving pattern is affected from this manipulation.

4.2 Experimental Design

To elicit conditionally altruistic preferences, each subject is paired with another subject and all of them are asked to make two choices: an unconditional choice (UC) and a conditional choice (CC). The unconditional choice stage is a standard strategy method version of the Dictator Game. It presents to both subjects the choice problem of a decision-maker in a standard Dictator Game. In this stage, out of the ten tokens they are given, subjects choose the number of tokens they wish to transfer to the other participant, keeping the rest for themselves. They know that one of the choices will be implemented. After they have completed this stage, they decide in the conditional choice stage how to allocate the same amount if they knew the choice of the transfer of their counterpart in the unconditional stage. Thus, in this stage subjects make 11 giving choices for each possible choice of their counterpart. Subjects in this stage are presented a table to state their conditional giving schedule, similar to the conditional cooperation schedule introduced by Fischbacher et al. (2001).[3]

Subjects in the lab decide as though they were assigned as the dictator, and they make their choices for both stages. Then the role of the subject and the

[3] See Section C.4 for the decision screens.

choice of the subject among UC and CC is randomly selected for the payment.

4.2.1 Treatments

The first treatment, which we will refer to as *Recipient ITA*, was conducted only in Trento, Italy, with the subject from the University of Trento's experimental economics lab (CEEL) pool. In first treatment, each subject was matched with another subject in the lab.

In the second treatment, which we will refer to as *Recipient UGA*, we paired subjects with individuals from a group that had lower socioeconomic status. This treatment was conducted with two parties in two locations. The first party, as in the RecipITA treatment, was a group of subjects from the pool of the experimental economics lab of University of Trento. The second party was chosen from the Achioli community, located in Kitgum District in Northern Uganda. Members of the Acholi community were earning around 2000-4000 Ugandan Schillings, which was equal to roughly 0.6 - 1.2 Euros for a day's work during the time the studies were conducted. The main income source of the members of the Achioli society was subsistence farming, and they did not have steady access to employment. Italian subjects were given this information in addition to some other information on the life standards and prices of general consumption goods of the participants in Uganda while a similar information about life standards of the Italian subjects was given to subjects from Uganda. The pictures in Figure 4.1 were shown to Italian subjects in order to increase salience on the socioeconomic status of the participants in Uganda. We assumed that Italian subjects already had a clear idea about the average socioeconomic status of their peers in *RecipITA* treatment; therefore, we did not provide similar information in this treatment. The sessions with Italian subjects were conducted with computers in the lab in both treatments.

The sessions with in Uganda were conducted with pen and paper, with the help of an experimenter and a translator.

Figure 4.1: A collage of pictures taken in the field, which show the participants in Uganda and their living environment. These pictures were shown to the Italian subjects in the *RecipUGA* treatment, in order to provide a salient picture of the socioeconomic status of the recipients.

We used a different procedure in order to eliminate complexities of conditional choice in the field in Uganda. In the *RecipITA* treatment, all subjects made decisions in the UC and CC stages we described. In the *RecipUGA* treatment, only the Italian subjects made decisions in both the UC and CC stages, while Ugandan subjects only made decisions in the UC stage. As we were only interested in the transfer by the Italian subjects towards Italians compared to their transfer to subjects in Uganda, we did not expect this change to affect our analysis.

Below we summarize the experimental flow:

RecipITA	*RecipUGA*
Instructions UC & CC	Instructions UC & CC (Italy)
Control questions	Control questions (Italy)
UC	UC (Italy)
CC	CC (Italy)
Coin tossing	Instructions UC (Italy)
Role assignment	Coin tossing (Italy)
Feedback and payment	UC (Uganda)
	Role assignment (Uganda)
	Feedback and payment (Uganda)
	Feedback and payment (Italy)

Table 4.1: Experimental Flow in Two Treatments

We ran five sessions for the *RecipITA* treatment in Trento, with 100 subjects in total; four sessions for the *RecipUGA* treatment in Trento, with 59 subjects in total; and three sessions with members of the Acholi Community in the Kitgum District in Uganda, with 59 participants in total. We used z-Tree for the lab sessions in Trento (Fischbacher, 2007).

4.2.2 Payoffs

In order to balance the payments according to purchasing power, we used different exchange rates for the subjects in Italy and subjects in Uganda. Subjects in Italy received 1 EUR (European Euro) for each token they earned, while subjects in Uganda earned 700 UGX (Ugandan Schilling) for each token, which corresponded to about 0.20 EUR at the time the study was conducted. These exchange rates were calculated according to the average daily income for both parties.

Selected Dictator	Selected Stage	*RecipITA*	*RecipUGA*	
			Italian	Ugandan
Player i	UC	$10 - UC_i$	$10 - UC_i$	$10 - UC_i$
	CC	$10 - (CC_i \vert UC_j)$	$10 - (CC_i \vert UC_j)$	-
Player j	UC	UC_j	UC_j	UC_j
	CC	$(CC_j \vert UC_i)$	-	$10 - (CC_i \vert UC_j)$

Table 4.2: Payoffs of Player i in Two Treatments. Please note that, if a subject in Uganda is selected as the dictator, his/her unconditional choice is implemented.

4.3 Results

Result 1 - Unconditional giving in *RecipUGA* is higher than it is in *RecipITA* treatment: Table 4.3 shows the summary statistics of the unconditional transfers made by subjects in Italy in two treatments. A Mann-Whitney-Wilcoxon(MWW) test reveals that the average unconditional transfer towards Italian subjects (2.05 tokens) is significantly lower than the average unconditional giving towards Ugandan subjects (4.05 tokens).

Treatment	Min	Median	Max	Mean	Std.Dev.	Wilcoxon p
RecipITA	0.00	2.00	6.00	2.06	1.78	$< .001$
RecipUGA	0.00	3.00	10.00	4.05	3.21	

Table 4.3: Comparison of Summary Statistics of Two Treatments

Figure 4.2 shows that the modal given amount in *RecipITA* treatment is 0 tokens, while the modal given amount in *RecipUGA* treatment is 3 tokens. In addition to that, transfers above 6 tokens are non-existent in the *RecipITA* treatment, whereas in *RecipUGA* treatment those constitute roughly 27% of all the transfers. In this treatment, around 14% of the participants decide to transfer the total pie to the participants with whom they have been matched. This result suggest that low socioeconomic status of recipients might increase the level of unconditional transfers.

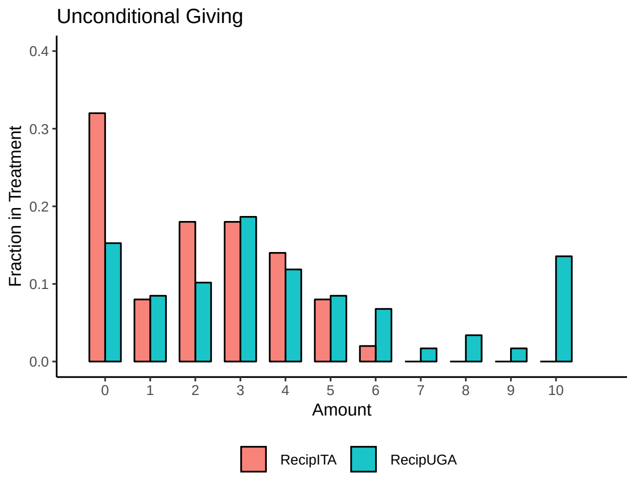

Figure 4.2: Distribution of Unconditional Transfers for Each Treatment

It might be argued that, as the subjects in Uganda in *RecipUGA* treatment make the unconditional giving choice, the difference between the two treatments in terms of unconditional giving might be attributed to this difference as well. This difference would potentially increase the importance of strategic component of the unconditional giving in the *RecipITA* treatment. However, we observe a similar distribution of unconditional giving in the *RecipITA* treatment compared to non-strategic dictator games in the literature. Moreover, the fact that a higher amount of giving is present in the *RecipUGA* treatment supports the hypothesis that the main difference is driven by altuistic tendencies. Moreover, the claim is supported by the comparision of unconditional and the conditional giving behavior of the subjects, which we discuss further in this section.

Result 2 - Conditional strategy distributions are similar in both treatments: To classify subjects according to their conditional behavior, we used the methodoloy by Fischbacher et al. (2001) which we described in the Appendix section C.2. . The decision patterns in the *CC* stage show

	Dependent variable:			
	Conditional	Selfish	Hump-Shaped	Unconditional
RecipUGA	0.520	-0.754^{*}	-1.111^{*}	0.373
	(0.331)	(0.445)	(0.659)	(0.291)
Constant	-0.282	-1.099^{***}	-1.815^{***}	-1.555^{***}
	(0.202)	(0.231)	(0.288)	(0.199)
Observations	159	159	159	159
Log Likelihood	-108.811	-79.650	-52.356	-44.185
Akaike Inf. Crit.	221.622	163.299	108.711	92.371
Note:			*p<0.1; **p<0.05; ***p<0.01	

Table 4.4: Logit Regression on Dummy Variable for Types.

a striking similarity in both treatments in terms of conditional behavior. Figure 4.3 shows the fraction of conditional types by treatment. Conditional altruists seem to be the most common type (43% for *RecipITA* and 55.6% for *RecipUGA*, and they are followed by selfish types (11.9% for each treatment). The order of unconditional subjects and hump-shaped subjects is reversed. In the *RecipUGA* treatment, the fraction of unconditional types is around 11.9% of the subjects, nearly double the fraction of unconditional cooperators, which is %6 in *RecipITA*. However, we cannot reject the hypothesis that the proportion of conditional cooperators is the same for both treatments, neither with Fischer's Exact Test (see Table C.1) nor with a logisric regression over whether the subject is unconditionally altruist or not (see Table 4.4). The logistic regression suggests that the proportions of selfish and hump-shaped individuals are both lower in *RecipUGA* treatment, however these differences are not significant.

Result 3 - In both treatments, the majority of subjects are conditional altruists: Whether giving is directed towards subjects with similar socioeconomic status (*RecipITA*), or towards subjects who have a lower socioeconomic status (*RecipUGA*), the majority of subjects are classified as conditional altruists. This result provides strong evidence for the presumptive reciprocity hypothesis.

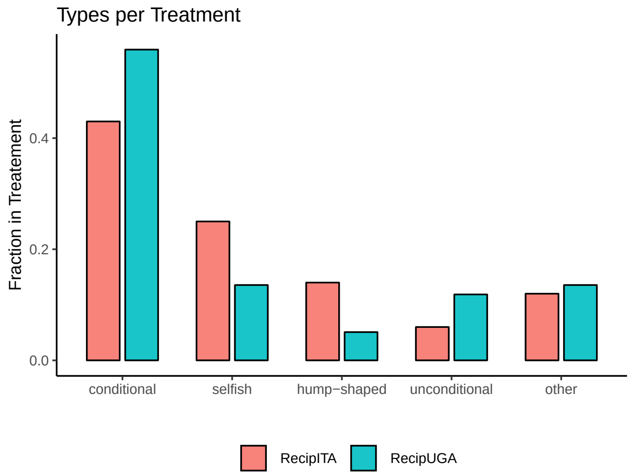

Figure 4.3: Type Classifications for Both Treatments According to the Decision in Conditional Choice Stage

We further look at the details concerning the behavior of conditional altruists. Our analysis is summarized in Table 4.5.

We compare two models which differ in terms of their assumptions on interaction effect between treatment and contribution of the other. According to Model 1, which assumes no interaction, in *RecipUGA* treatment subjects give 0.90 tokens unconditionally while for each token the other would give we observe a 0.57 token increase to the given amount. In Model 2, which assumes the interaction effect, a unit increase in the other subject leads to an increase of 0.53 tokens in both treatments. Moreover, it seems that conditionality of giving is even stronger in the *RecipUGA* treatment, which leads to a further nearly 0.09 tokens more. the Akaike Information Criteria (AIC) favors Model 1, while the Bayesian Information Criteria (BIC) favors Model 2. Regardless of which model to take into account, we conclude that the higer need of the opponent does not decrease the slope of the conditional giving function.

Dependent variable: Conditional Giving		
	Model 1	Model 2
(Intercept)	0.09	0.28
	(0.24)	(0.24)
UC_{other}	0.57***	0.53***
	(0.01)	(0.02)
RecipUGA	0.90**	0.46
	(0.35)	(0.37)
UC_{other}:RecipUGA		0.09***
		(0.03)
AIC	2926.50	2922.95
BIC	2950.14	2951.32
Log Likelihood	-1458.25	-1455.48
Num. obs.	836	836
Num. groups: subject	76	76
Var: subject (Intercept)	2.10	2.10
Var: Residual	1.47	1.45

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4.5: Linear Mixed Model Regression Analysis of the Conditional Giving for the Types that are Classified as Conditional Altruists. Decisions are clustered for each subject.

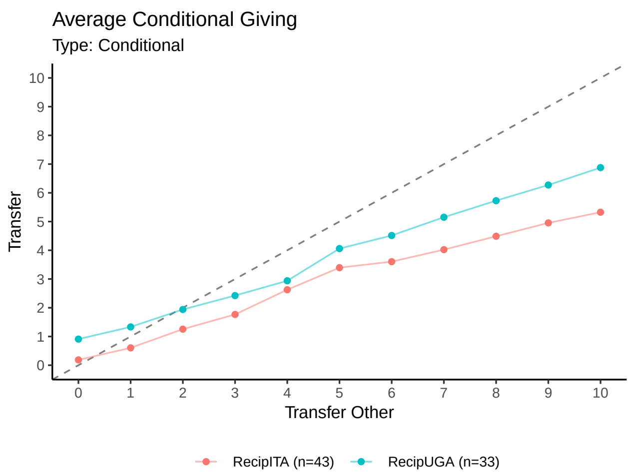

Figure 4.4: Average Conditional Choice of the Subjects who are Classified as Conditional Altruists

Result 4 - The difference between treatments in UC is mostly caused by unconditional altruists Considering that our results suggest that the most common types are conditional altruists, and the distribution of conditional strategies is not different in two treatments, the puzzle emerges as to how to explain the difference in unconditional giving between two treatments. At first glance, this would seem to contradict with our main hypothesis on presumptive reciprocity, as it is counter-intuitive to expect a higher level of giving from those who have a lower socioeconomic status. To explain this behavior, we look at the unconditional giving of the subjects with different conditional types.

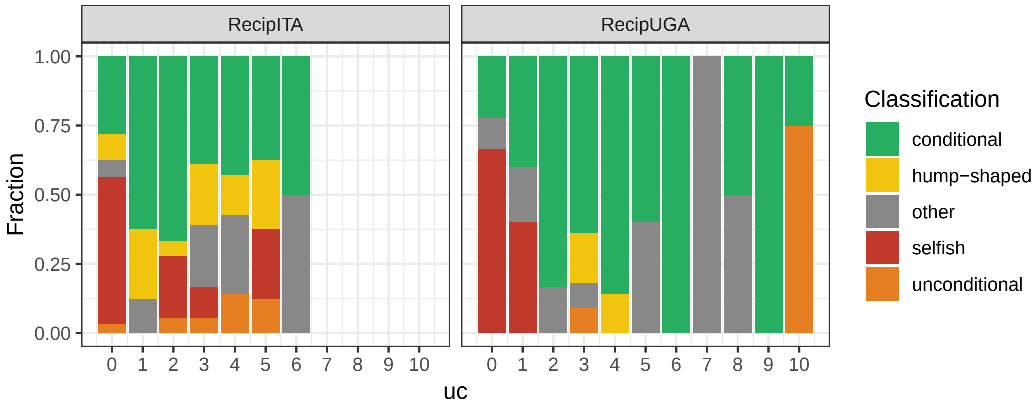

Figure 4.5: Unconditional Choices and the Classification of Donors According to Their Types in the Second Stage

Figure 4.5 shows the proportion of conditional types in the second stage grouped by their unconditional choice in the first stage. In the *RecipITA* treatment, unconditional altruists transfer on average around 3 tokens to the recipient, while in the *RecipUGA* treatment a vast majority of those types transfer the maximum possible amount ten tokens, which results in an average of nine tokens. Their conditional giving pattern is aligned with the first-stage behavior: Figure 4.6 shows the average conditional strategy of unconditional types in the second stage.

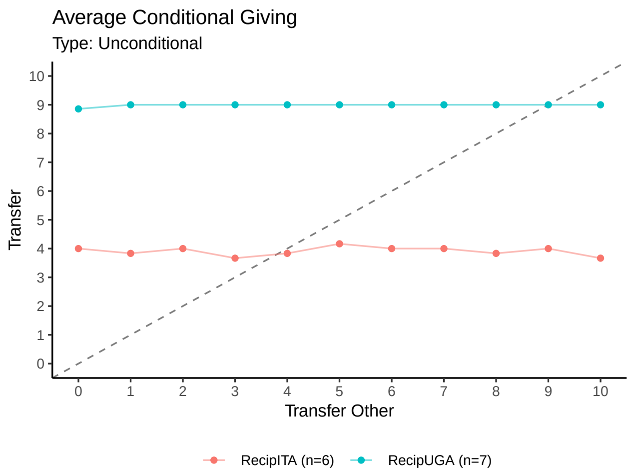

Figure 4.6: Average Conditional Choice of the Subjects who are Classified as Unconditional Altruists.

To illustrate this further, we analyze unconditional giving with a Generalized Linear Model shown in Table 4.6. In this model, we analyze unconditional choices, with respect to conditional types and treatments. The first model ignores the interaction effect of the two dependent variables, while the second model, which explains our data better according to the Akaike information criterion and the Bayesian information criterion, considers it. In the second model, the increase in unconditional giving is explained jointly by the treatment effect which has the size of 1.821 tokens and the interaction between unconditional type and the treatment effect which has the size of 4.179 tokens. That amount is more than double the base treatment effect. Obviously, it would be a mistake to assume that the same participants would be unconditional altruists in both treatments. However, as the distribution of types is similar, the evidence suggests that the main driver of such differences in the dictator game are those who have unconditional giving preferences.

Overview of Conditional Giving: Before we further discuss the implications of our results, we demonstrate aggregated conditional strategies of all types. Figure 4.7 shows a similar slope of both treatments, while

Dependent variable: Unconditional Giving		
	Model 1	Model 2
(Intercept)	2.31***	2.21***
	(0.29)	(0.30)
hump-shaped	−0.00	0.22
	(0.57)	(0.61)
other	0.35	0.71
	(0.53)	(0.64)
selfish	−1.91***	−1.25*
	(0.44)	(0.49)
unconditional	3.07***	0.79
	(0.63)	(0.86)
RecipUGA	1.59***	1.82***
	(0.35)	(0.46)
hump-shaped:RecipUGA		−0.92
		(1.33)
other:RecipUGA		−0.86
		(1.01)
selfish:RecipUGA		−2.53**
		(0.92)
unconditional:RecipUGA		4.18***
		(1.19)
AIC	695.20	678.01
BIC	716.68	711.77
Log Likelihood	-340.60	-328.00
Deviance	675.40	576.43
Num. obs.	159	159

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4.6: Generalized Linear Mixed Model (GLMM) Regression for the Dependent Variable Unconditional Choice.

having different intercepts which mostly caused by the increased giving of unconditional altruists.

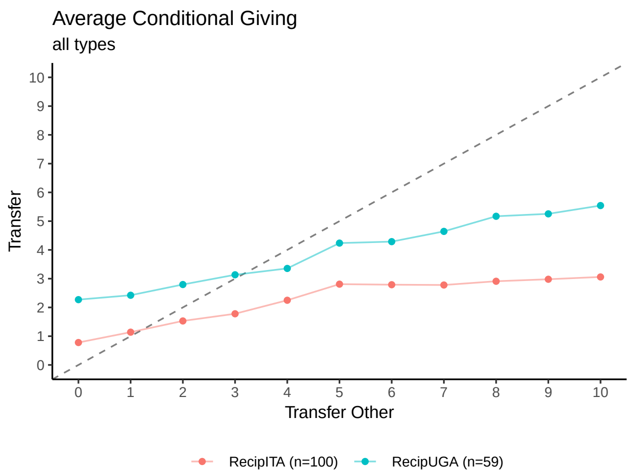

Figure 4.7: Average Conditional Giving

To investigate this pattern, we run a Generalized Linear Mixed Model (GLMM) regression of which results are shown in Table 4.7. The best model to explain the data following AIC and BIC, is Model 2, which confirms our graphical demonstration of the results. According to the model, the effect of low socioeconomic status has two implications in overall subjects: it increases the intercept by nearly one token, and it rewards the generosity of the opponent with an extra 0.12 tokens, in addition to the mean conditional marginal reward of 0.23 tokens.

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	0.87***	1.10***	−0.29	−0.10	0.14
	(0.23)	(0.23)	(1.34)	(1.34)	(1.38)
UC_{other}	0.28***	0.23***	0.23***	0.23***	0.23***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
RecipUGA	1.66***	1.06**	1.02**	1.02**	0.98*
	(0.36)	(0.39)	(0.39)	(0.38)	(0.39)
UC_{other}:RecipUGA		0.12***	0.12***	0.12***	0.12***
		(0.03)	(0.03)	(0.03)	(0.03)
Age			0.06	0.07	0.07
			(0.06)	(0.06)	(0.06)
Male				−0.64	−0.62
				(0.35)	(0.35)
Econ Student					−0.26
					(0.37)
AIC	7200.59	7185.92	7190.61	7189.52	7191.16
BIC	7227.92	7218.72	7228.88	7233.26	7240.36
Log Likelihood	-3595.30	-3586.96	-3588.31	-3586.76	-3586.58
Num. obs.	1749	1749	1749	1749	1749
Num. groups: subject_unq	159	159	159	159	159
Var: subject_unq (Intercept)	4.65	4.66	4.65	4.58	4.60
Var: Residual	2.71	2.68	2.68	2.68	2.68

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4.7: Linear Mixed Model Regression Analysis of Conditional Giving.

Overall, our results show that conditional giving is the main driver of the giving in dictator games. We do not observe any age effects, gender effects, or economics education.

4.4 Discussion

In this study, we explored whether the giving in the Dictator Game can be explained by a form of reciprocity based on the presumed kindness of the other person by collecting preferences of conditional giving to the kindness of the recipient. Moreover, we aimed to manipulate this relationship in a treatment where the recipients have low socioeconomic status.

Overall, our results show that conditional altruism dictates giving patterns when subjects can condition their giving to the kindness of the recipient, counter-intuitively also towards the low socioeconomic status. Although the intercept of conditional giving are different both treatments, conditional giving is strong in both, and even stronger in the setting in which the recipient has a higher degree of need.

We can think of two arguments against our conclusion from the results we obtained; first, our design does not eliminate strategic giving in the first stage (UC). One might expect subjects to tend to give more to the recipients if they expect recipients to employ conditionally altruistic preferences (Engelmann and Fischbacher, 2009; Vu, 2018). Moreover, sophisticated subjects thus infer strategic intentions of their counterpart and and can adjust their conditional giving accordingly. As we compare giving in UC with the previous literature, we can conclude that the former claim is unlikely. Average giving and the distribution of given amounts are similar to the previous experiments. Although we cannot be certain about the second claim, if indeed some of the subjects are able to do that kind of reasoning, the shift would create is expected to be towards a selfish type. Thus, if this mechanism is in action, we might only underestimate the number of conditional types and the slope of conditional decisions.

The second argument is about the moral wiggle room: some subjects can exploit the possibility of conditional decisions by using the conditional choice as a tool to reduce their giving for some conditional decision levels that look socially appropriate. Therefore, if they reduce their giving conditioned to small number of tokens given by others, they might look as if they are conditional altruists. Although there is no direct way for us to check the intention behind the conditional preferences, we can expect that an exploiter's unconditional giving should also be equal to or greater than the subject's maximum conditional giving in the second stage. In other words, they should not be willing to give more than their unconditional choice. Indeed , we observe that indeed three out of 43 conditional altruists in the *RecipITA* treatment, and four out of 33 conditional altruists in the *RecipUGA* treatment employ such a pattern. Their exclusion would not affect our results.

In our design, we used strategy method to elicit the unconditional and the conditional preferences. The evidence on whether subjects state their preferences differently in strategy method compared to direct direct behavior is mixed (Brandts and Charness, 2011). Even though, the evidence suggests that the behavior in stategy method and the direct elicitation of responses are similar in many games Public Goods Game (Fischbacher et al., 2012), the use of strategy method might have an effect that we cannot disentangle with the current design we have especially since the characteristics of the non-strategic dictator game and the Public Good Game differs. Moreover, both the dictator and the recipient role. Burks et al. (2003) suggests that playing both roles in might have two consequences for our design in different directions: the increase of emphaty which may lead to an increase in the the giving, and, the decrase of responsability which might lead to a decrease in giving. They show that in the Trust Game the latter hypothesis is stronger as the average contributions decline their contribution when they play the both roles overall. On the other hand, Iriberri and Rey-Biel (2011) shows that playing both roles in a mini-Dictator-game-like distribution game leads to a higher level of altruism. We acknowledge this fact might affect our two treatments in different directions as well and have an affect that might interfere with the results.

Chapter 5

Conclusion

This dissertation has investigated the role of reciprocity on altruistic and cooperative preferences. It has been known for a long time that reciprocity is a strong mechanism that can foster and sustain cooperation. Moreover, the experimental economics literature shows that the majority of people employ conditional strategies in common social dilemma games; thus, understanding the dynamics of reciprocity carries substantial importance to foster cooperation.

A common pattern in Public Goods Game experiments is that cooperation declines over time when the game is played repeatedly. The decline in cooperation is often attributed to *selfish-biased conditional cooperation* claims that this decline we observe is due to the dynamics of conditional strategies, an argument based on the assumption that conditional strategies revealed in lab experiments are stable. Chapter 2 - *On the Stability of Conditional Cooperation* - tested the validity of this assumption and found that it is not a correct assumption: We show that, as the game is played repeatedly, conditional cooperation declines and the fraction of selfish subjects increases remarkably. Our results show that changing beliefs and learning over time affects the conditional preferences of the individuals. This has been the case whether the conditional preferences of the opponent are visible or not. Therefore, imitation effects seem to be weak. These results have two main implications. The first one is a methodological one: It might be misleading to use conditional strategy as a measure for cooperative tendencies, as is frequently done. The second hints at the sustainability of cooperation: in the course of repeated interactions, even if people employ perfectly conditional strategies at the beginning, once cooperation is collapsed it would be more difficult to refoster cooperation by individual efforts as reciprocal preferences are influenced as well.

One key component of cooperation and reciprocity relationship has transpired to be repetition: if individuals are likely to interact repeatedly, it is possible for cooperation to emerge given that most of the people employ reciprocal preferences. Our study *The Evolution of Conditional Cooperation*, we presented in Chapter 3, creates a link between the experimental economics literature on conditional cooperation and the interdisciplinary literature on the evolution of cooperation which employs theoretical and computational methods. By using simulations, we estimated how the likelihood of cooperation emerging changes over different continuation probabilities, and which conditional types are likely to emerge. This creates a more direct comparison between the behavior of subjects in experimental games such as the Prisoners' Dilemma, the Public Goods Game and the Trust Game. Our results showed a significant parallels between the strategies that emerged in our simulations and the types we observe in the literature. In the study, moreover, we provided a rationale for the "hump-shaped" type of strategies that were commonly observed in experiments, but, due to its seemingly counter-intuitive structure, lacked a sound explanation.

Finally, in *Presumptive Reciprocity in Dictator Games* in Chapter 4 we investigate the discrepancy between giving in dictator games and giving in real life. We try to elicit the motives behind the altruistic behavior we observe in dictator games by collecting conditional giving strategies. Our results reveal that most of the subjects employ conditional giving preferences: They tend to give less or nothing if they know that the recipient would give nothing in the same position, and they increase their giving if the opponent gives more of the pie in the game regardless whether the other opponent is in need or not. These results call the "pure altruism" in question. We show that revealed altruism has strong parallels to the contribution preferences in the Public Goods Game. Although this is surprising, in terms of cooperation it does not draw a negative picture. As pure altruists are more open to exploitation, in a mixed population, it gives an advantage to selfish types.

These three studies we presented in this dissertation provide insights for cooperation problems in real-life, especially when formal institutions which can support cooperation are not available or building those instiutions that can monitor individual efforts are not possible or feasible. Our overall results point out to the the strong tendency of conditionality in both cooperation and altruitic behavior. Considering the potential erosion of conditional cooperation suggest that individuals adapt the situations when cooperation is not beneficial anymore. If majority of people are endowed with such conditional preferences, it might be easier to quickly obtain cooperation due to reciprocity. But the combined effect of strong conditionality of certain preference types (such as all-or-none type of conditional cooperatators) and the unstability of these preferences might lead to a quick collapse of cooperation in which individual efforts afterwards may not suffice to resurrect due to the erosion of conditionality as well. Future research will reveal to what extent this instability of conditionality is driven by the specific setting of our own experiment and clarify in what conditional conditional prefences can be stabilized.

Appendices

Appendix A

Supplementary Information "Stability of Conditional Cooperation"

A.1 Classification of Conditional Types

We denote each conditional strategy by three letters that represent the conditional responses to L, M, and H, respectively. Since there are only three responses, in total 27 different conditional strategies are possible. As the conditional strategies in our game are simple, the conditional type classification is rather self-evident; thus, we classify conditional types without the need of a calculation or a subjective evaluation.

We say that a player is selfish if he/she maximizes his/her own payoff (LLL); *conditional cooperator* if he/she increases its contribution monotonically. A special case of this type of strategy is the *perfect conditional cooperator* (LMH), who responds to the counterpart with the same action the counterpart played. We refer to conditional cooperators who are not perfect conditional cooperators as *imperfect conditional cooperators*[1]. We call a strategy *hump-shaped* if the most generous response of the player is a response to M (LML, LHL, LHM, MHL, MHM). We classified conditional strategies that do not fit any of the definitions above as *other* type. Figure A.1 demonstrates the conditional strategies graphically and their classifications.

[1]Please note that we also classify some of the generous strategies such as MMH and MHH as “imperfect” since they do not perfectly copy the opponents’ actions. In that sense, it differs from the “selfish-biased” meaning. In practice, however, this do not make any difference, since no subjects use either of these imperfectly generous strategies.

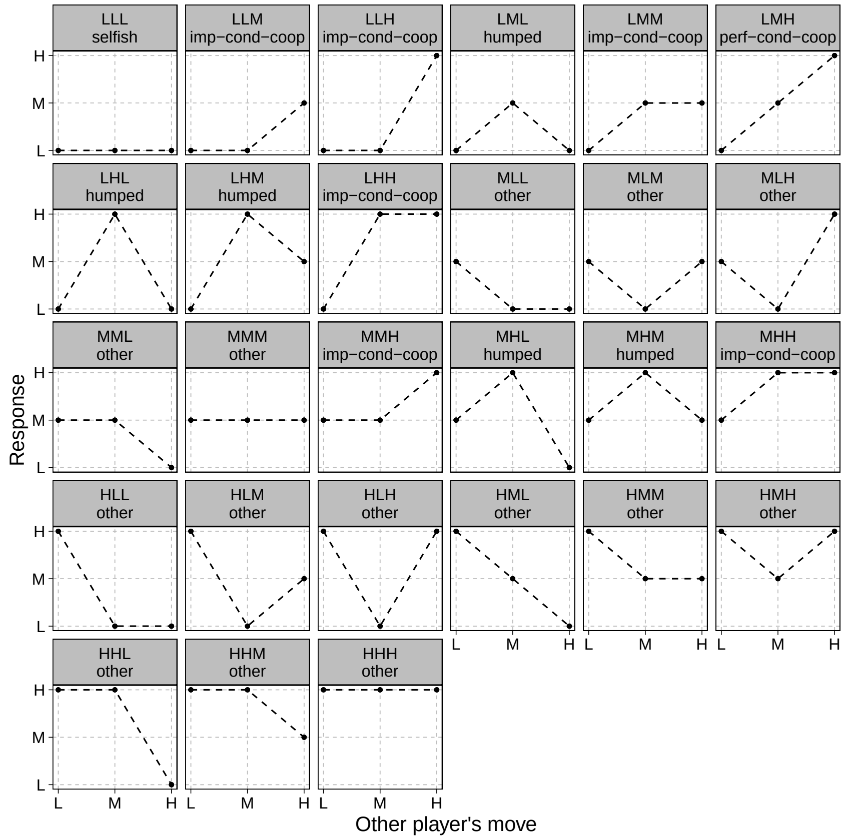

Figure A.1: All Possible Conditional Types and Their Classifications

A.2 Belief Elicitation Task

We used the Quadratic Scoring RuleBrier, 1950a to elicit subjects' beliefs on the opponents' action in an incentive-compatible manner. To increase the comprehension, we used an alternative presentation of the Quadratic Scoring Rule Artinger et al., 2010. We designed an interface which contains a slider for each possible outcome to inform subjects about their possible earnings from the belief elicitation task. We also provided a table in the printed instructions which shows the relevant reward and penalty for the probability assigned to an action.

The reward of a subject from one choice in the belief elicitation task in which a subject revealed his/her guess as $p = (p_L, p_M, p_H)$ and action $j \in$

$\{L, M, H\}$ is chosen by the opponent, denoted $Q_A(p)$, and calculated as:

$$Q_j(p) = \alpha + \beta - \beta(1 - p_j)^2 - \beta \sum_{i \neq j} (p_i)^2, \tag{A.1}$$

We used parameters $\alpha = 20$ and $\beta = 20$ in our experiment, which give each subject the possibility to earn between 0 Experimental Currency (ECU) and 40 ECU in the belief elicitation task stage. For instance, if a subject assigns a 100% probability to the action chosen by the counterpart, he/she receives 40 ECU as a reward and 0 tokens as a penalty, which adds up to 40 ECU. Similarly, if a subject assigns a 50% probability to the action chosen by the counterpart and a 25% probability for the other two actions, he receives 35 ECU as reward and 5 ECU as a penalty for each of the other alternatives that are not selected, which adds up to 25 tokens received from the belief task for this particular action.

Assigned Probability(%)	*Reward if chosen (EC)*	*Cost if not chosen (EC)*
100	40	20
95	39.95	18.05
90	39.8	16.2
85	39.55	14.45
80	39.2	12.8
75	38.75	11.25
70	38.2	9.8
65	37.55	8.45
60	36.8	7.2
55	35.95	6.05
50	35	5
45	33.95	4.05
40	32.8	3.2
35	31.55	2.45
30	30.2	1.8
25	28.75	1.25
20	27.2	0.8
15	25.55	0.45
10	23.8	0.2
5	21.95	0.05
0	20	0

Table A.1: Rewards and Costs of Correct Beliefs according to Quadratic Scoring Rule

A.3 Results on Beliefs

(Brier, 1950a). We concentrate on the two most interesting cases, in which beliefs concern the first player's unconditional choice (Figure A.2) and the second player's choice conditional to the first player choosing an high transfer (Figure A.3). In both pictures, the gray bars at the bottom represent the difference between actions and beliefs and the green line is the regression line. Figure A.2 reveals that, on average, subjects' beliefs about the first players' choices are remarkably accurate in the initial rounds and show no tendency of becoming more or less accurate in later rounds. Figure A.5 reveals that in the first stages of the game subjects are also clever at guessing the second player's transfer in response to an high transfer. Interestingly, however, their guesses become *less* accurate as the game unfolds, as they become *more* (unconditionally) selfish than they think the other subjects are. Notice that this lends further support to the thesis that the decline of conditional cooperation is mostly driven by conditionally co-operative subjects switching to pure selfishness.

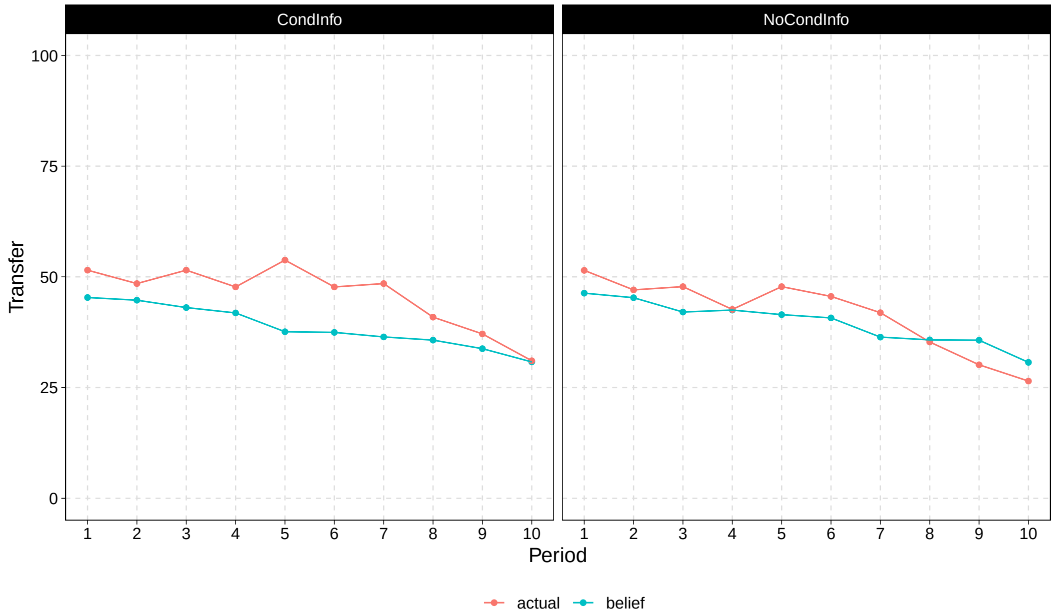

Figure A.2: Mean Unconditional Transfer and Beliefs on Expected Transfers in Each Treatment.

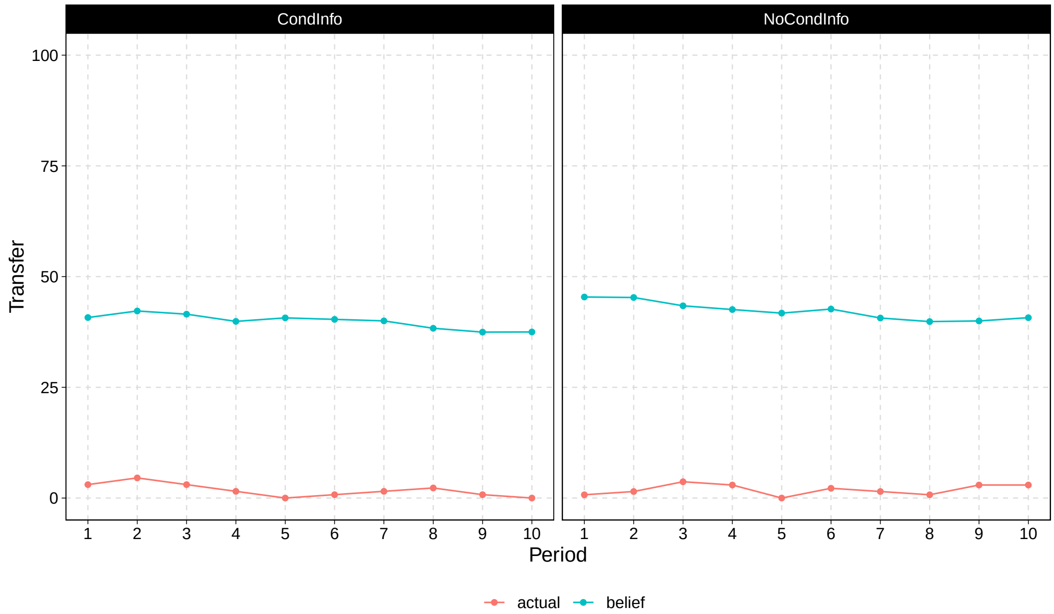

Figure A.3: Mean Conditional Transfers and Beliefs on Expected Transfers in Response to L in Each Treatment.

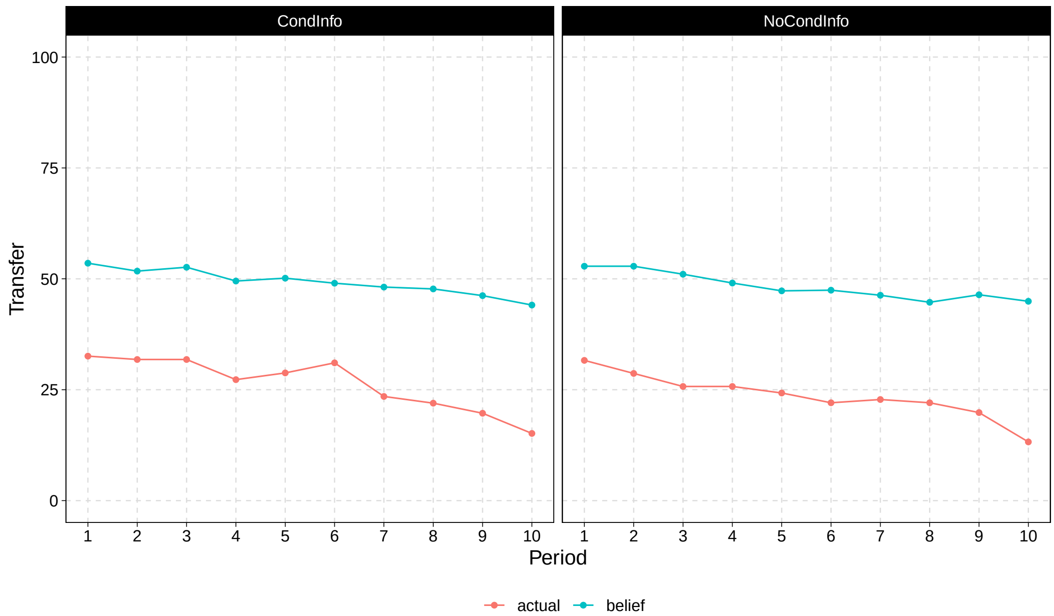

Figure A.4: Mean Conditional Transfers and Beliefs on Expected Transfers in Response to M in Each Treatment.

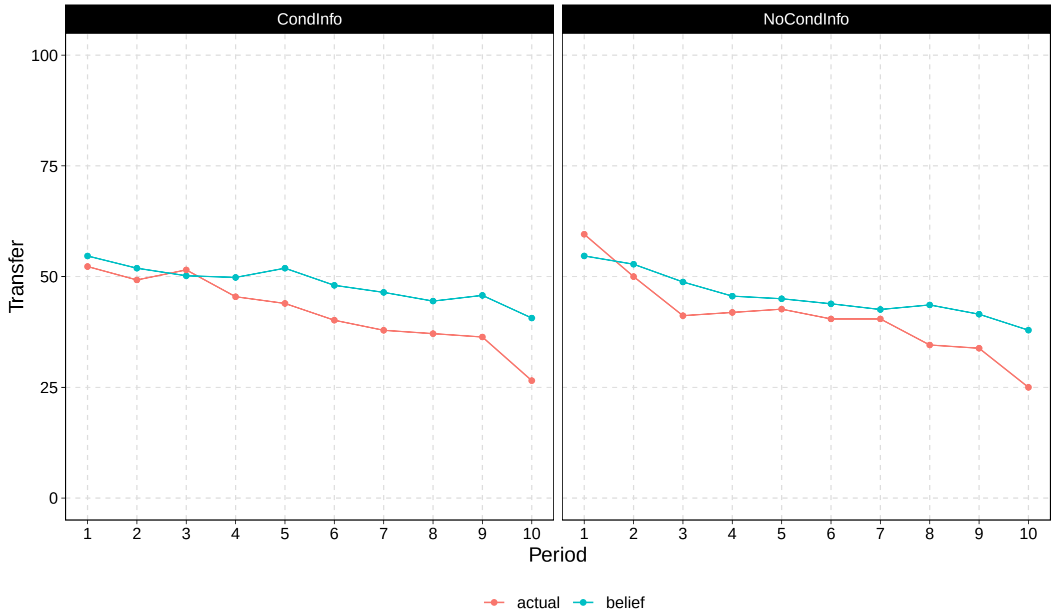

Figure A.5: Mean Conditional Transfers and Beliefs on Expected Transfers in Response to H in Each Treatment.

A.4 Comparision of Proportions of Conditional Types with Fischbacher(2001)

	FGF2001	Per. 1	Per. 2	Per. 3	Per. 4	Per. 5	Per. 6	Per. 7	Per. 8	Per. 9	Per. 10
Cond. Coop.	22	87	82	75	71	75	69	68	63	61	46
Selfish	13	35	39	47	53	46	50	55	60	65	79
Humped	6	10	8	8	8	13	12	9	9	5	7
Other	1	2	5	4	2	0	3	2	2	3	2
p-value		0.412	0.333	0.383	0.313	0.262	0.739	0.368	0.266	0.041	0.011

Table A.2: Comparision of Number of Subjects for Each Type. p-values in each column show the results of comparision of proportions with the Fischbacher et. al. (2001) study by using Chi-Squared Test of Equal Proportions

A.5 Comparision of Proportions of Conditional and Selfish Types

Period	*CondInfo*	*NoCondInfo*	p-value
1	0.62	0.68	0.62
2	0.61	0.62	1.00
3	0.62	0.50	0.22
4	0.53	0.53	1.00
5	0.56	0.56	1.00
6	0.47	0.56	0.39
7	0.48	0.53	0.73
8	0.47	0.47	1.00
9	0.50	0.41	0.39
10	0.36	0.32	0.76

Table A.3: Proportions of ConditioNal Cooperators in Two Treatments and p-value of Chi-squared Test of Equal Proportions

Period	*CondInfo*	*NoCondInfo*	p-value
1	0.29	0.24	0.62
2	0.29	0.29	1.00
3	0.30	0.40	0.34
4	0.39	0.40	1.00
5	0.32	0.37	0.67
6	0.35	0.40	0.69
7	0.41	0.41	1.00
8	0.44	0.46	0.99
9	0.45	0.51	0.60
10	0.56	0.62	0.62

Table A.4: Proportion of Selfish Players in Two Treatments and p-value of Chi-squared Test of Equal Proportions

A.6 Transaction Rates Between Types

Table A.5 shows the fraction of transactions from one conditoinal strategy to another during ten periods.

	LLL	LLM	LLH	LML	LMM	LMH	LHL	LHM	LHH	MLL	MLM	MML	MMM	MMH	MHM	MHH	HML	HHH
LLL	0.85	0.03	0.00	0.02	0.03	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
LLM	0.17	0.63	0.02	0.01	0.12	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00
LLH	0.18	0.09	0.64	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LML	0.13	0.01	0.01	0.62	0.10	0.10	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
LMM	0.18	0.14	0.00	0.04	0.51	0.12	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00
LMH	0.11	0.02	0.02	0.02	0.10	0.69	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
LHL	0.00	0.00	0.00	0.00	0.12	0.00	0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12
LHM	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LHH	0.00	0.00	0.00	0.00	0.00	0.36	0.09	0.00	0.45	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00
MLL	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MLM	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MML	0.40	0.00	0.00	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00
MMM	0.40	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00
MMH	0.00	0.00	0.00	0.00	0.00	0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.00	0.00
MHM	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00
MHH	0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.14	0.00	0.43	0.00	0.00
HML	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.17	0.00
HHH	0.33	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00

Table A.5: Transastion Probabilities Calculated as the Average Transaction Frequency in Ten Periods

A.7 Screenshots

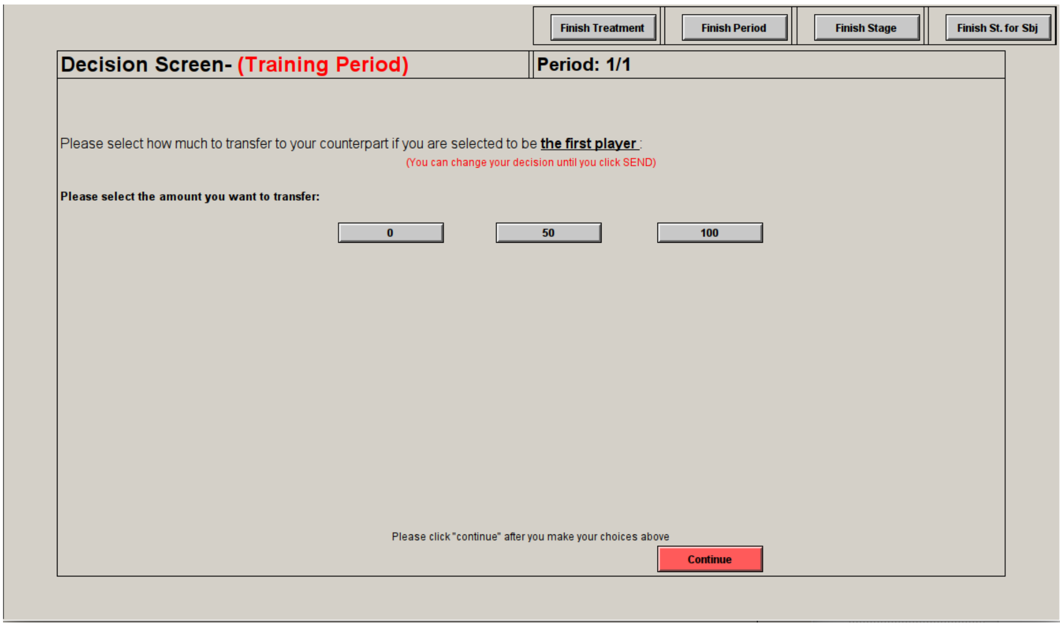

Figure A.6: Transfer Decision Screen as the First Player

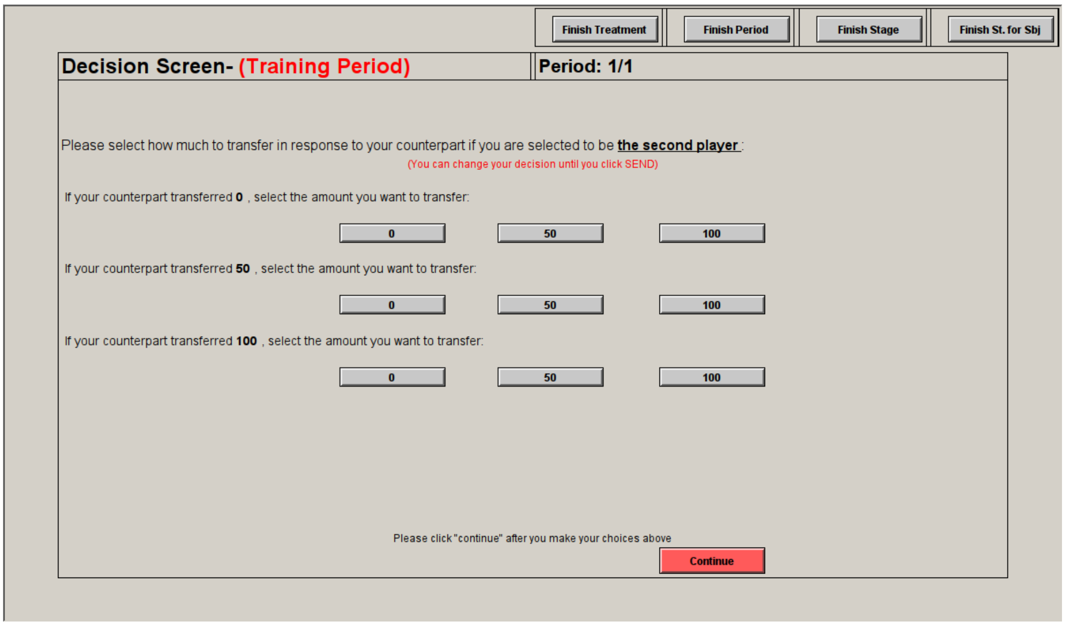

Figure A.7: Transfer Decision Screen as the Second Player

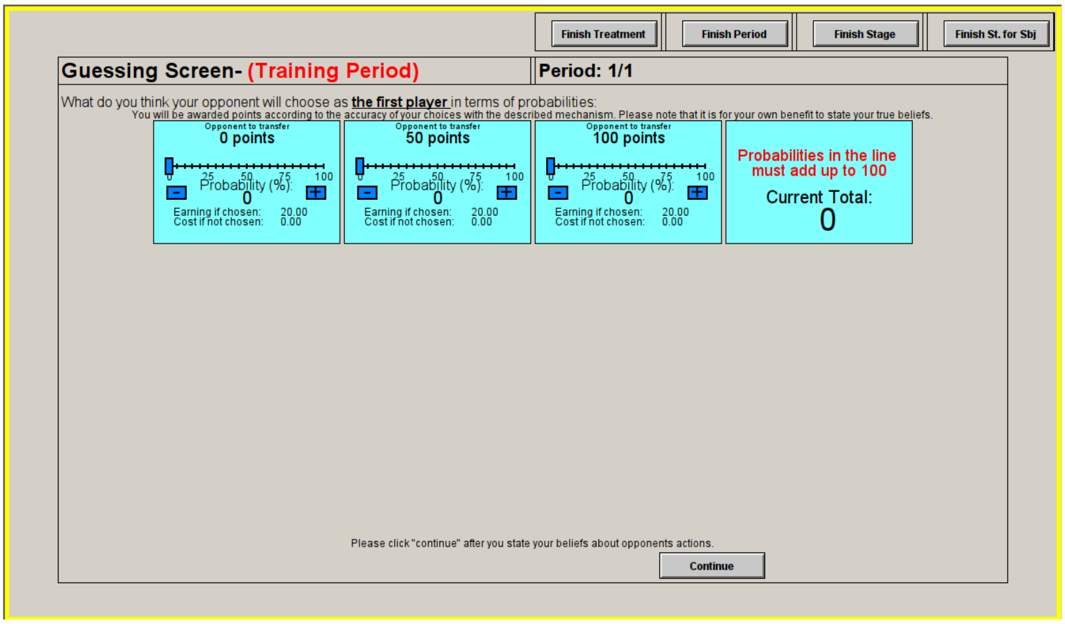

Figure A.8: Belief Elicitation Screen as the First Player

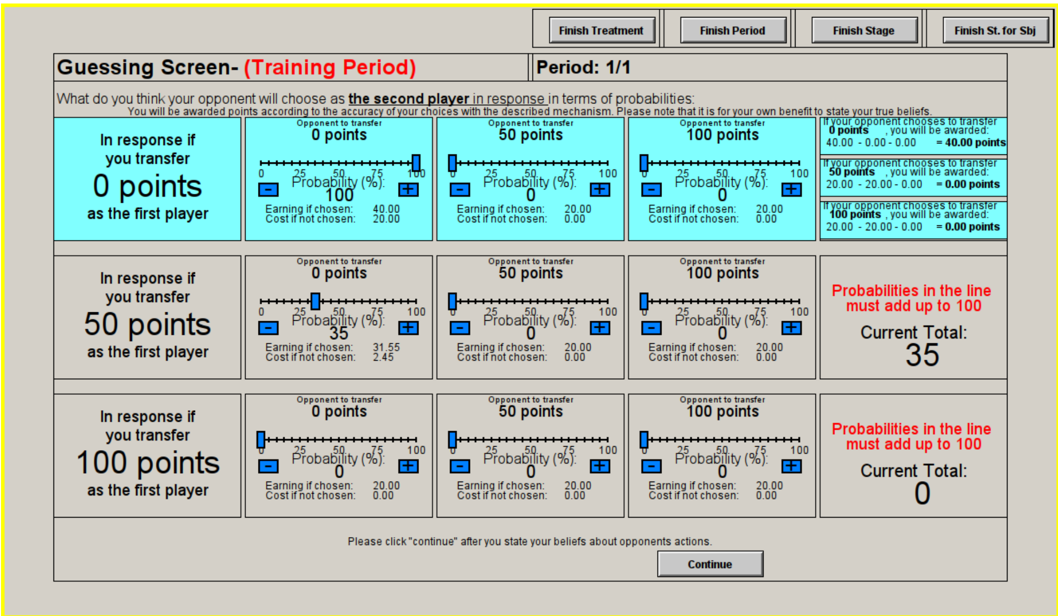

Figure A.9: Belief Elicitation Screen as the Second Player

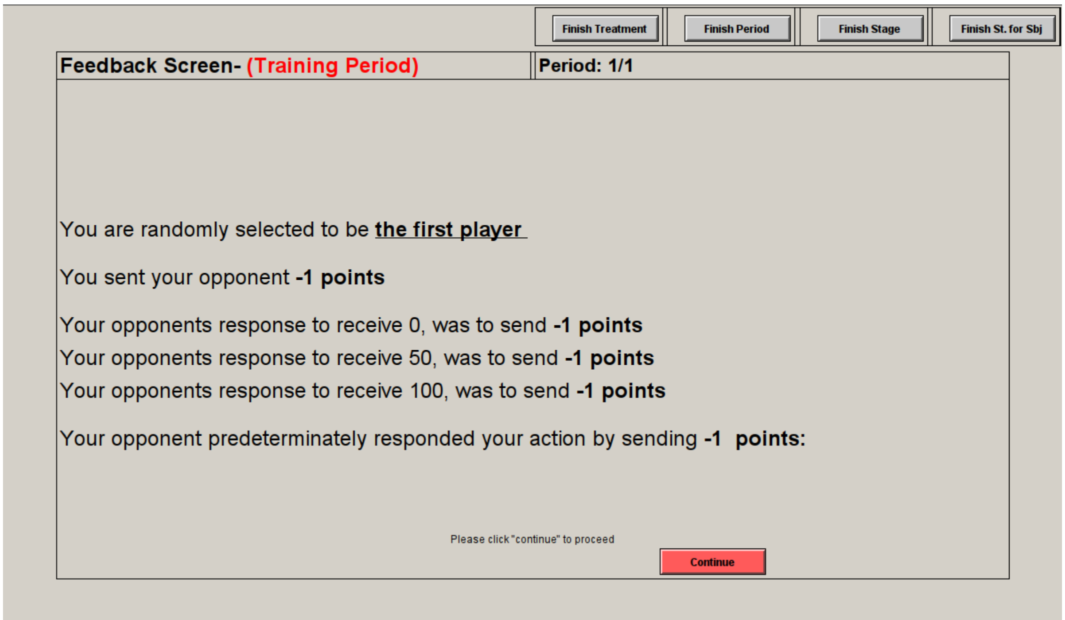

Figure A.10: Feedback Screen for the First Player. The three lines shows the conditional strategy of the opponent starting from the third line only appears in *CondInfo* treatment.

A.8 Experimental Instructions

Instructions

General Information

Welcome to the experiment and thank you for your participation. This experiment is funded by **University of Trento** and all the data acquired will be used for scientific purposes.

From now on please remain silent, do not communicate with other participants and raise your hand if you have questions. Using cell phones and any other means of communication are completely prohibited. Please note that violators of these rules will be excluded from both the experiment itself and all ensuing payments.
In this experiment, depending on the decisions you and the other participants make, you can earn money.

The payment will be made after the experiment with Euros in cash. You will be given **3 EUR** by default for your participation and additional to that amount you might earn more money depending on the procedure will be explained here. However during the experiment all the currency will be in **"Experimental Currency (EC)"**. After the experiment, your earnings in EC will be converted in Euros with the following conversion rate:

1 EC = 0.05 EUROS

You will be interacting with other participants in some parts of the experiments. During and after the experiment, your and other participants decisions will remain anonymous. I.E., the identity of you and the participant(s) you will interact will be kept secret. The payment you will earn from the experiment will be given personally and will not be revealed to other participants. Moreover all the data acquired from the experiment will be kept without the identities of the participants.

Structure

The experiment will consist 10 repeated rounds identical to each other and each period will consist two stages that you will make decisions: **Interaction Stage** and **Guessing Stage**. Prior to 10 identical periods of these two stages, there will be a 4 rounds of **Training Stage** where you will practice the experiment and **Questionnaire Stage** posterior to it where you will be asked to give feedback about the experiment and provide some demographic questions.

The general structure of the experiment can be summarized by the table:

<table>
<tr><th>Stage Name</th><th>Repetition</th></tr>
<tr><td>Training Stage</td><td>x4</td></tr>
<tr><td>Guessing Stage</td><td rowspan="2">x10</td></tr>
<tr><td>Interaction Stage</td></tr>
<tr><td>Questionnaire</td><td></td></tr>
<tr><td>Payments</td><td></td></tr>
</table>

Procedure

Interaction Stage

In this stage you will be matched with a random participant in the lab anonymously. You will have an initial endowment of **100 EC** each round and you will face some transfer decision to the opponent as it will be described.

One of the players will play as the **first player,** and the other player will be the **second player. However y*ou will make your decisions both as if you will be selected as the first player and the second player each round.*** *After you and your partner make the decisions for first and second player role,.* the roles will be assigned and the decisions will be executed accordingly.

The first player will have the option to transfer **0**, **50** or **100 EC** to the opponent out of his 100 EC and keep the rest for himself/herself. The amount that the first player transfer will be multiply by **3.** That is;if the first player selects zero to transfer, he will keep all 100 EC to himself/herself and the opponent will get nothing, if he/she selects to transfer **50 EC** to the opponent, he will keep the **50 EC** to himself the opponent will receive **150 EC.** And finally if he decides to transfer **100 EC,** he will keep nothing and the opponent will receive **300 EC.** The decision screen of the first player will be shown in the screen as the following:
[First Player Decision Screenshot Here]

The second player will have the option to transfer also **0, 50** or **100 EC** out of 100 EC to the opponent and as the same above and similarly the transfer that the second player made to the first player will be multiplied by **3**. However the second player will have the opportunity to choose his/her transfer to the opponent, conditioned what the first player chose to transfer to himself/herself.

So the second player will have three decisions:
If the other player chooses **0 EC ,** to transfer **0, 50** or **100** EC;
If the other player chooses **50 EC ,** to transfer **0, 50** or **100** EC;
If the other player chooses **100 EC ,** to transfer **0, 50** or **100** EC.

And the actual transfer will be made according to what the first player transfers to second player.

The choice problem of the second player will be shown in the screen as the following:

[Second Player Decision Screenshot Here]

As both players have 100 EC and the amount decided to be transferred will be multiples by three, the summary of earnings for both players will be as the following table:

		Second Players' Response to first players given choice		
		0 EC	**50 EC**	**100 EC**
First Players' Transfer Choice	**0 EC**	First Player earns 100 EC, (100 kept, 0 received) Second Player earns 100 EC. (100 kept, 0 received)	First Player earns 250 EC, (100 kept,150 received) Second Player earns 50 EC. (50 kept, 0 received)	First Player earns 400 EC, (100 kept,300 received) Second Player earns 0 EC. (0 kept, 0 received)
	50 EC	First Player earns 50 EC, (50 kept, 0 received) Second Player earns 250 EC. (100 kept, 150 received)	First Player earns 200 EC, (50 kept, 150 received) Second Player earns 200 EC. (50 kept, 150 received)	First Player earns 350 EC, (50 kept, 300 received) Second Player earns 150 EC. (0 kept, 150 received)
	100 EC	First Player earns 0 EC, (0 kept, 0 received) Second Player earns 400 EC. (100 kept, 300 received)	First Player earns 150 EC, (0 kept, 150 received) Second Player earns 350 EC. (50 kept, 300 received)	First Player earns 300 EC, (0 kept, 300 received) Second Player earns 300 EC. (0 kept, 300 received)

After you have selected your decisions as the both first and the second player, the computer will select either you or your opponent as the first player and the other as the second player with equal probability. The first players choice and the second players corresponding response will be taken to realize the interaction and determine the earnings in that round.

Note that each period you have the equal chance to be the first player and the second player. Therefore it is probable that in some rounds you will be selected as the first player, in some you will be selected as the second player.

The interaction will be repeated 10 times and **in each period you will be matched with a different player in the lab (You will never be matched with the same person again)**.

At the end of the experiment,one of the periods will be selected and the earnings from that period will be taken for the actual payments.

Feedback about the Interaction

[TREATMENT NoInfo]: After each period there will be feedback about opponents decisions: If you are selected to be the first player, **you will see your opponents corresponding response to your choice as the second player**, and your opponent will **see your choice as the first player.** If you are selected to be the second player, **you will only see your opponent's choice as the first player** and your opponent will see **your corresponding response as the second player**. An example screen about the feedback screen is like the following:

[Treatment ACT Feedback Screenshot Here]

[TREATMENT Info] : After each period there will be feedback about opponents decisions: If you are selected to be the first player, you will see **your opponents conditional choices for each choice of yours**, and **your opponent will see your choice as the first player**. Similarly if you are selected to be the second player, you will only see your **opponent's choice as the first player** and your opponent will see **your conditional choices for each possible choice of him/her** as the second player. An example feedback screen is like the following:

[Treatment CST Feedback Screenshot Here]

Summary of the Interaction Stage

- Both players select their amount of transfer as the first player (0 , 50 or 100 EC).
- Both players select their choice as the second player (0, 50, or 100 EC conditioned on each three possible transfers of the first player).
- One player will be selected as the first player, one player will be selected as the second player randomly by the computer.
- The player who is selected as the first player realizes his transfer, the player who is selected to be the second player will realize his transfer conditioned on what the other player choose.)
- They will receive a feedback about the opponents decision.

Guessing Stage

Before you and your opponent make a decision in each Interaction Stage, you will be asked to assess probabilities based on your opponents decisions for the following Interaction Stage. Your earning from this part will be based on your accuracy of your assessment of opponents decisions in the first stage.

You will be asked about your beliefs about your opponent will decide in the interaction stage in terms of probabilities (percentages sum up to 100). Since in each decision , players have three options (to transfer 0 EC, 50 EC or 100 EC) , you will be asked to put a probability for each option that your opponent have . One for the opponents choice as the first player, You can allocate the probabilities by both using the sliders (dragging the blue handle to the left and right) or plus and minus signs in each field.

Since in the *Interaction Stage*, each player will make four decisions, you will be asked to state your beliefs in four different choices: One for opponents choice as the first player in the first screen, three for your opponents choices in the second screen in the second screen.

You will see a screen as shown below to state your beliefs about your opponent's decision as the first player:

[First Player Guessing Screenshot Here]

After completing that screen you as shown below too state your beliefs about your opponent's decision as the second player:

[Second Player Guessing Screenshot Here]

For each decision your opponent have, you will earn an amount of money for the probability you assigned for the actual choice your opponent have. The more probability you put to a right choice, the more EC you will earn as reward (values are shown on the second column of the table below). However if you put an amount of probability to a decision that is not chosen, you will incur some cost. The more probability you put to a wrong choice, the more EC it will be reduced as the cost (values are shown on the third column of the table below)

Since there are three options for each choice, to transfer 0, 50 or 100 EC, you will get the reward for the right choice, while you will have the cost of other two choices will be reduced from your earning. The rewards and the costs you will have for a choice is the same for all choices and for all decision problems. The table for rewards and costs for the probability assigned to a choice are shown below:

Assigned Probability to an option (%)	**Reward if this option is chosen (EC)**	**Cost if this option is not chosen (EC)**
100	40	20
95	39.95	18.05
90	39.8	16.2
85	39.55	14.45
80	39.2	12.8
75	38.75	11.25
70	38.2	9.8
65	37.55	8.45
60	36.8	7.2
55	35.95	6.05
50	35	5
45	33.95	4.05
40	32.8	3.2
35	31.55	2.45
30	30.2	1.8
25	28.75	1.25
20	27.2	0.8
15	25.55	0.45
10	23.8	0.2
5	21.95	0.05
0	20	0

On the screens where you make your choice, you will be able to see the rewards and the costs for the probability you asses for each choice and the total earnings for each possibility. However if you prefer, you can refer the table during the experiment.

EXAMPLE:
Assume that you stated that the probability of your opponent to choose to transfer 50 EC is 80% , probability to choose to transfer 0 EC is 10% and probability to choose 100 EC is 10% as well for a decision he has to make. If your opponent actually chooses to transfer 50 for that decision;

You will receive the the corresponding reward for 80 percent for the right choice, that is, 39.2 EC
You will be reduced the corresponding costs for 10 percent for the two wrong choices, that is 0.2 EC for each.
Your earning will be 39.2 – 0.2 – 0.2 = 38.8 EC.

EXAMPLE:
Assume that you stated that the probability of your opponent to choose to transfer 0 EC is 50% , probability to choose to transfer 50 EC is 30% and probability to choose 100 EC is 20%. If your opponent actually chooses to transfer 50 EC for that decision ;

You will receive the the corresponding reward for 30% for the right choice, that is, 30.2 EC
You will be reduced the corresponding cost for 50% for the first wrong choice, that is 5 EC
You will be reduced the corresponding cost for 20% for the first wrong choice, that is 0.8 EC

Your earning will be 30.2 – 5 – 0.8 = 24.4 EC.

It can be seen that the maximum amount you can earn for your choice is 40 EC when you assign 100% probability to an option chosen (40 – 0 – 0) and minimum amount that you can get is 0 EC assigning 100% to an option that is not chosen (20 – 20 – 0).

Please note that with this mechanism it is for your benefit to reflect your choices according to your true beliefs.

At the end of the experiment, from all the guesses you make, one of the guess will be selected about the opponents choices and you will be paid according to accuracy that guess. There will be no feedback after guessing stages and the amount you will earn from the chosen guess will be notified at the end of the experiment and will be added to your payment.

Summary of the Guessing Stage

- Both players will state their beliefs according to what the opponent will choose to do as the first player and as the second player.
- They will be rewarded for their assigned probability for the opponent actual choice and they will have cost for their assigned probability for the opponent incorrect choice for each of the four decision problem each round.
- There will be no feedback after the guessing stage.

Training Stage

For you to get familiar to the Interaction Stage and Guessing Stage, there will be a training stage for 4 rounds. In this stage instead of being matched with participants in the lab, computer will make the decisions randomly as your opponent. You will not earn or lose any money for that stages. Please note that the computer will not make any calculations but will make his choices randomly.

In the first two rounds you will only go to the **Interaction Stage** and for the other two rounds you will see the **Guessing Screens** first and **Interactions Screens** afterwards as if you are playing with a real person.

As noted after the 4 rounds of Training Stage, you will start the actual experiment interacting with other players for 10 stages.

We are arrived at the end of the instructions. Please raise your hand if you have questions regarding the experiment.

Appendix B

Supplementary Information "Evolution of Conditional Cooperation"

B.1 Type Definitions

	denotion	classification			denotion	classification
0	L-LLL	selfish		41	M-MMH	conditional
1	L-LLM	conditional		42	M-MHL	humped
2	L-LLH	conditional		43	M-MHM	humped
3	L-LML	humped		44	M-MHH	conditional
4	L-LMM	conditional		45	M-HLL	other
5	L-LMH	perf-conditional		46	M-HLM	other
6	L-LHL	humped		47	M-HLH	other
7	L-LHM	humped		48	M-HML	other
8	L-LHH	conditional		49	M-HMM	other
9	L-MLL	other		50	M-HMH	other
10	L-MLM	other		51	M-HHL	other
11	L-MLH	other		52	M-HHM	unconditional
12	L-MML	other		53	M-HHH	unconditional
13	L-MMM	unconditional		54	H-LLL	selfish
14	L-MMH	conditional		55	H-LLM	conditional
15	L-MHL	humped		56	H-LLH	conditional
16	L-MHM	humped		57	H-LML	humped
17	L-MHH	conditional		58	H-LMM	conditional
18	L-HLL	other		59	H-LMH	perf-conditional
19	L-HLM	other		60	H-LHL	humped
20	L-HLH	other		61	H-LHM	humped
21	L-HML	other		62	H-LHH	conditional
22	L-HMM	other		63	H-MLL	other
23	L-HMH	other		64	H-MLM	other
24	L-HHL	other		65	H-MLH	other
25	L-HHM	unconditional		66	H-MML	other
26	L-HHH	unconditional		67	H-MMM	unconditional
27	M-LLL	selfish		68	H-MMH	conditional
28	M-LLM	conditional		69	H-MHL	humped
29	M-LLH	conditional		70	H-MHM	humped
30	M-LML	humped		71	H-MHH	conditional
31	M-LMM	conditional		72	H-HLL	other
32	M-LMH	perf-conditional		73	H-HLM	other
33	M-LHL	humped		74	H-HLH	other
34	M-LHM	humped		75	H-HML	other
35	M-LHH	conditional		76	H-HMM	other
36	M-MLL	other		77	H-HMH	other
37	M-MLM	other		78	H-HHL	other
38	M-MLH	other		79	H-HHM	unconditional
39	M-MML	other		80	H-HHH	unconditional
40	M-MMM	unconditional				

Table B.1: All Possible Types and Their Classifications

B.2 Additional Data

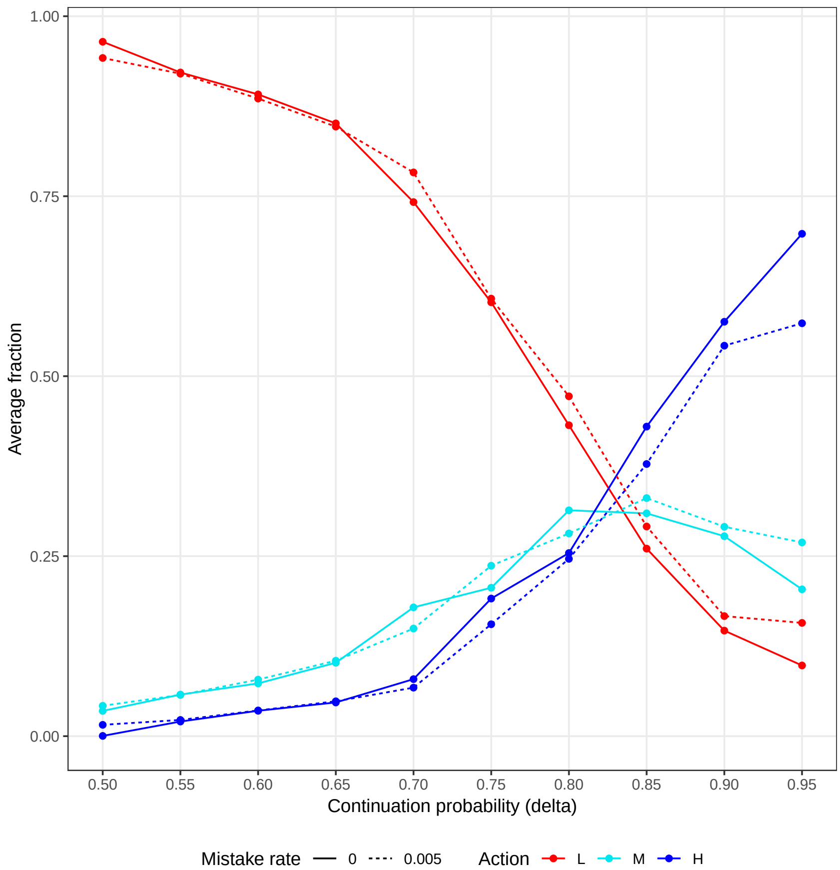

Figure B.1: Comparison of Actions with and without mistakes.

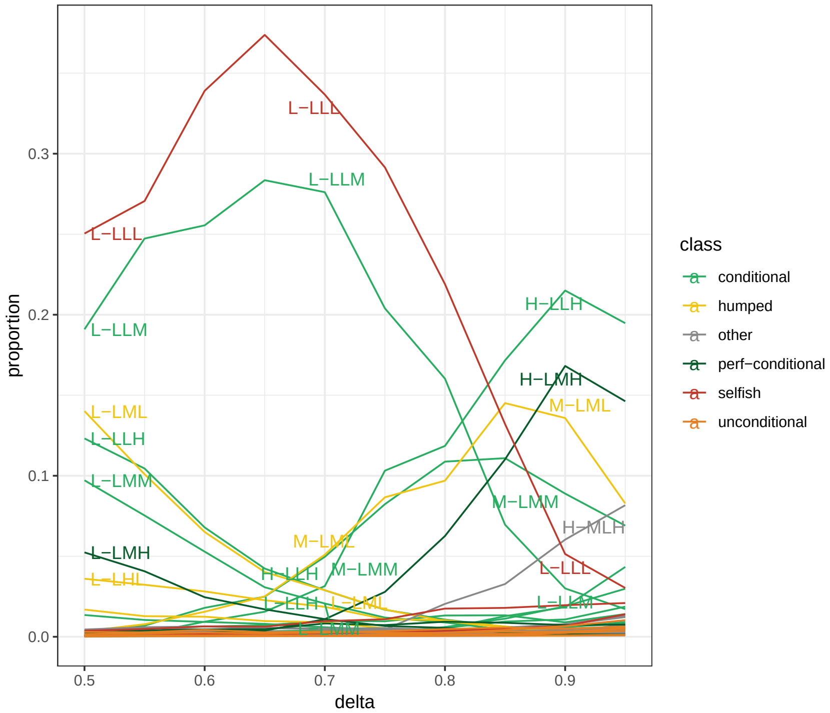

Figure B.2: All types' performance for different continuation probabilities.

Appendix C

Supplementary Information: Presumptive Reciprocity in Dictator Games

C.1 Additional Data and Analysis

C.1.1 Distribution of Types

Treatment		Conditional	Selfish	Unconditional	Hump-shaped	Other	Total
RecipITA	n	43	25	6	14	12	100
	%	43.00	25.00	6.00	14.00	12.00	
RecipUGA	n	33	8	7	3	8	59
	%	55.93	13.56	11.86	5.08	13.56	
p value		0.1397	0.1063	0.2349	0.1105	0.8074	

Table C.1: Type distributions in treatments and Fisher's test for differences in proportion. Results show no treatment effects in the distribution of types

C.1.2 Unconditional Giving Uganda

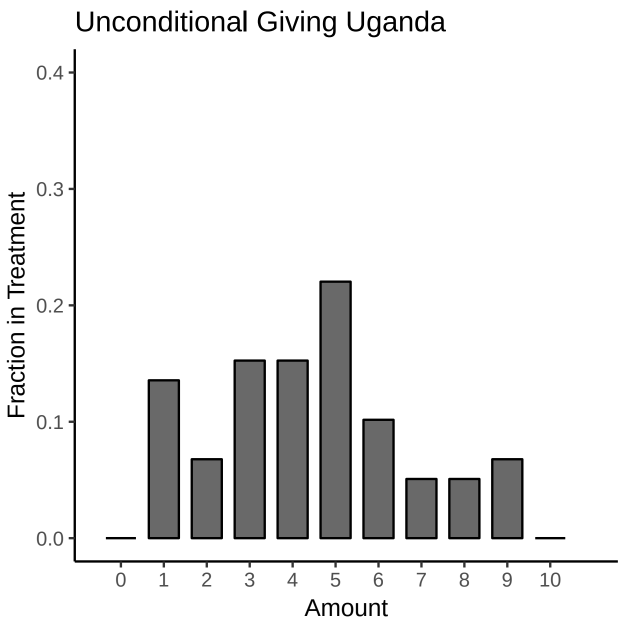

Figure C.1: Distribution of Unconditional Givng by Ugandan Subjects

C.2 Classification of Types

To classify subjects we use a similar procedure to (Fischbacher et al., 2001). We classified a subject *conditional altruist* if Spearman rank correlation coefficient between other's giving and the response to it is greater than zero and significant at 0.05 level. We classified a subject *unconditional altruist* if the correlation is insignificant, standard deviation is smaller than one and average contribution is greater than or equal to one; *selfish* if the average contribution is smaller than one. We identified *hump-shaped* givers visiually. And we classified the rest of the subjects as *other*. Following section all subjects' unconditional and conditional giving patterns.

C.3 Individual Giving Schedules Categorized by Type

Treatment: Recipient ITA

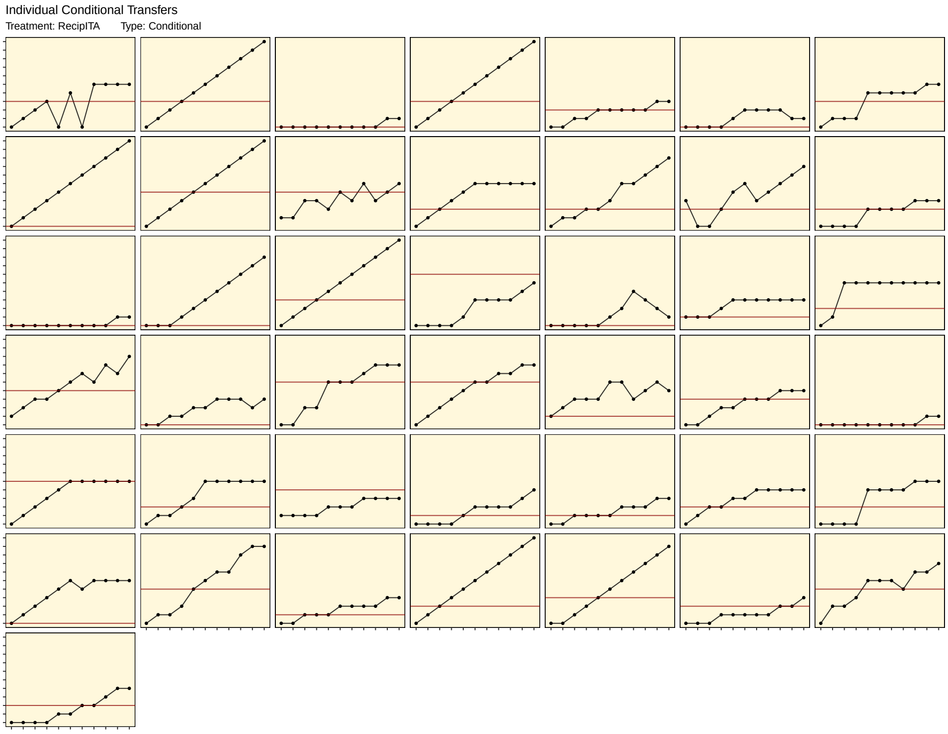

Figure C.2: Giving Schedule: Conditional Types

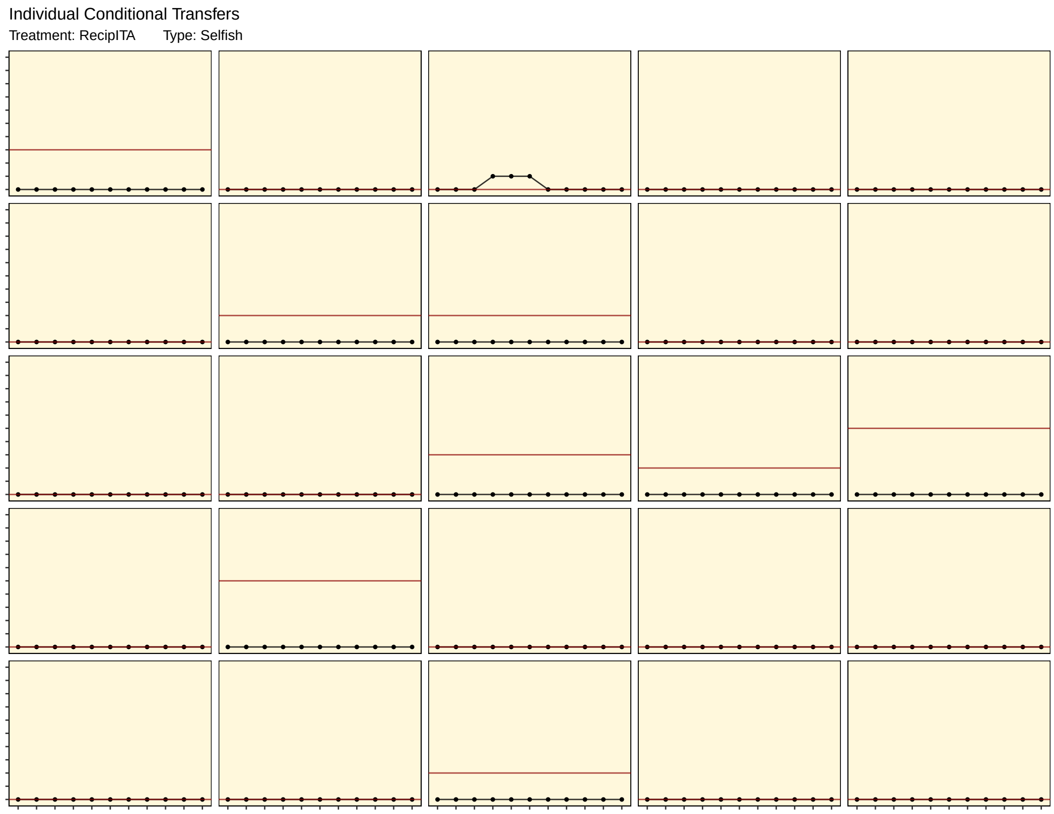

Figure C.3: Giving Schedule: Selfish Types

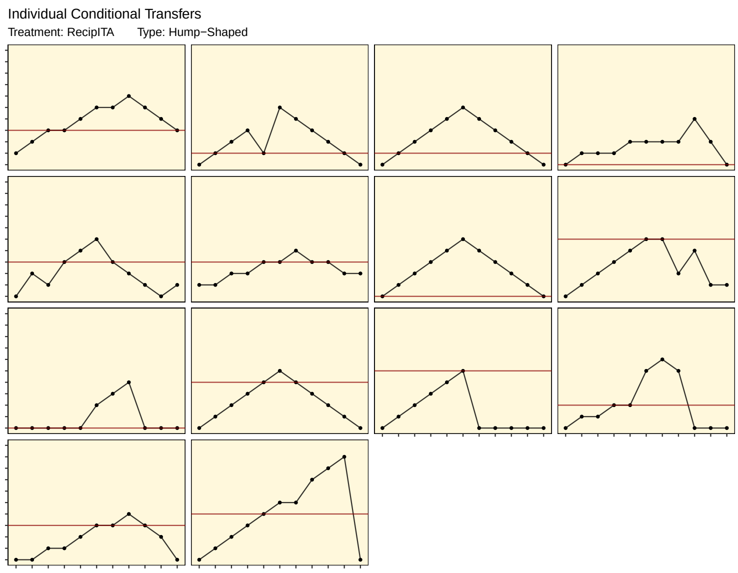

Figure C.4: Giving Schedule: Hump-Shaped Types

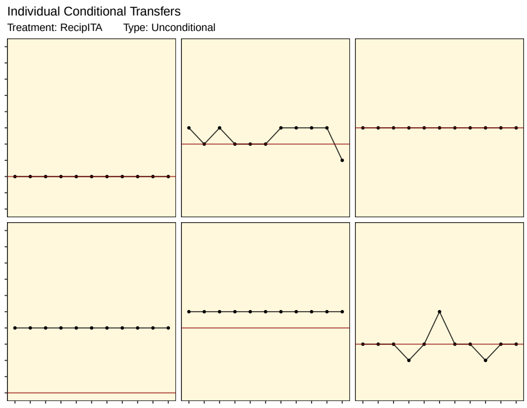

Figure C.5: Giving Schedule: Unconditional Types

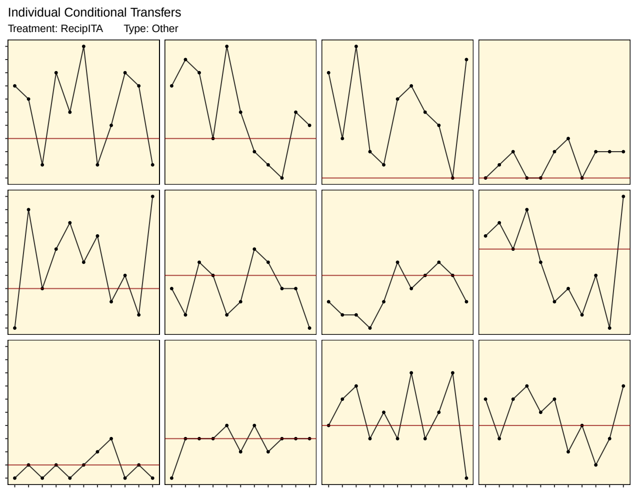

Figure C.6: Giving Schedule: Unclassified Types

Treatment: Recipient UGA

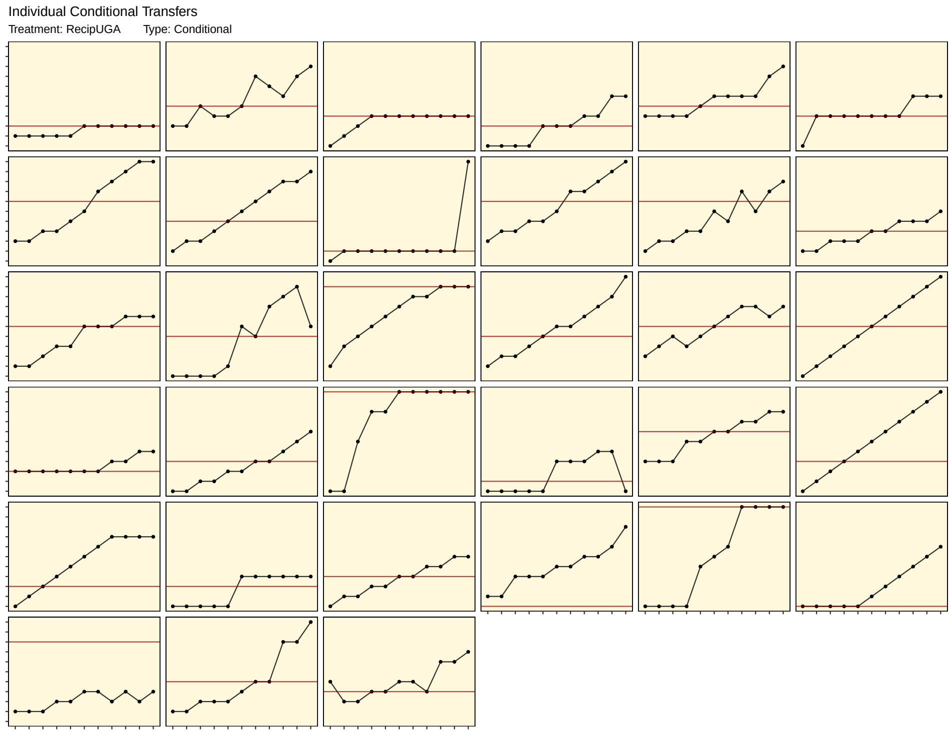

Figure C.7: Giving Schedule: Conditional Types

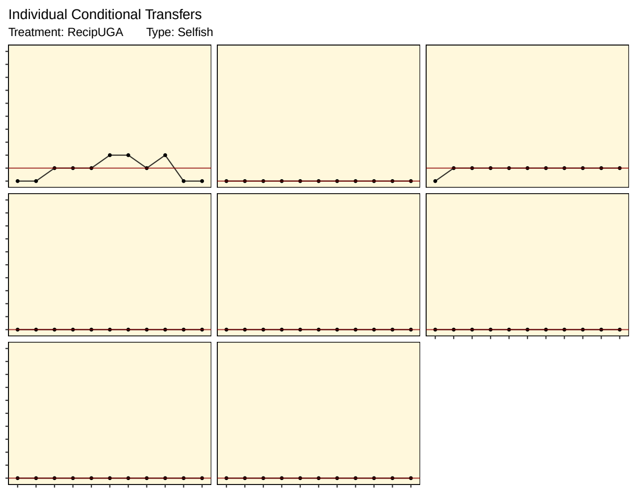

Figure C.8: Giving Schedule: Selfish Types

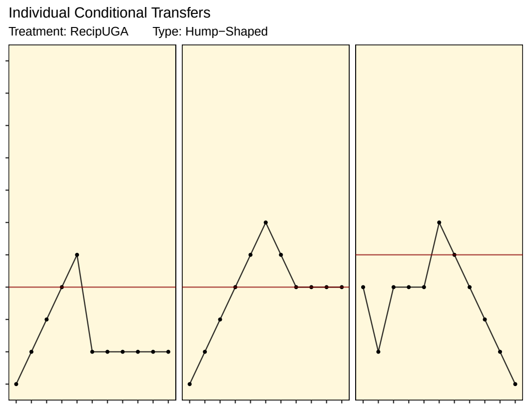

Figure C.9: Giving Schedule: Hump-Shaped Types

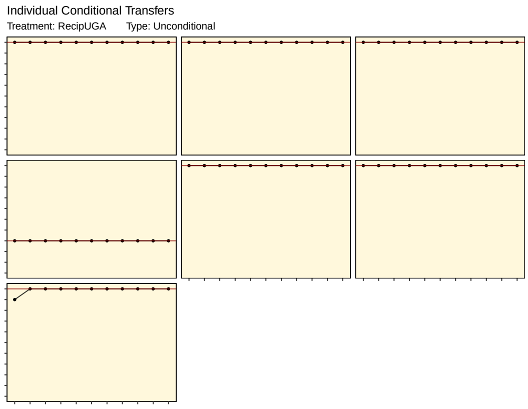

Figure C.10: Giving Schedule: Unconditional Types

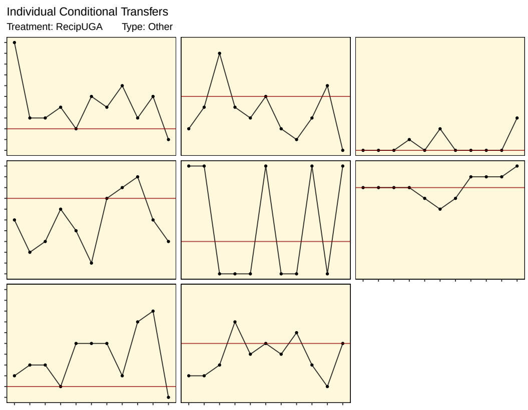

Figure C.11: Giving Schedule: Unclassified Types

C.4 Screenshots for Italian Participants

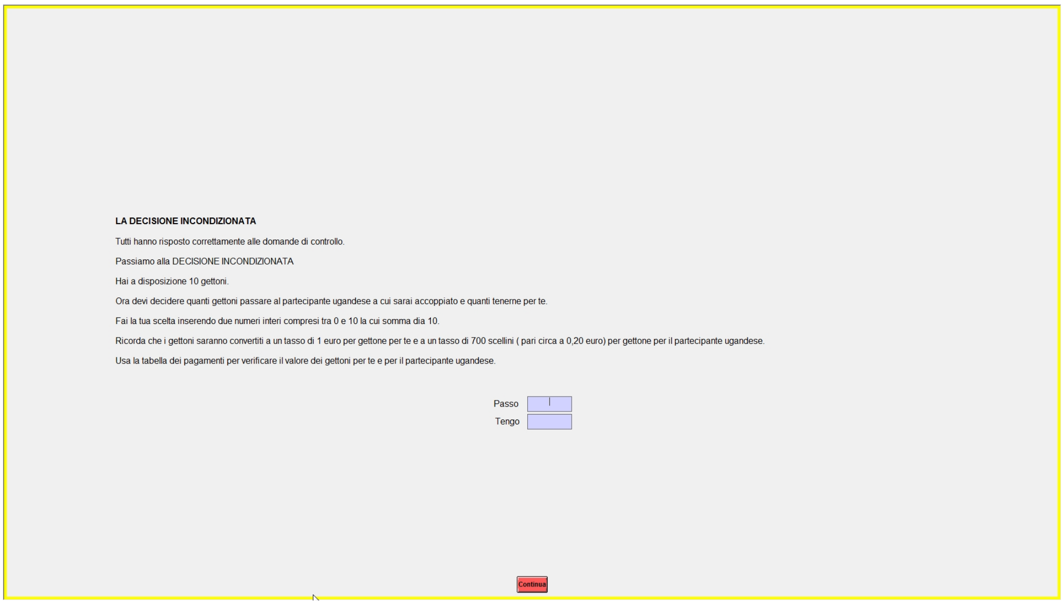

Figure C.12: Decision Screen: Unconditional Choice

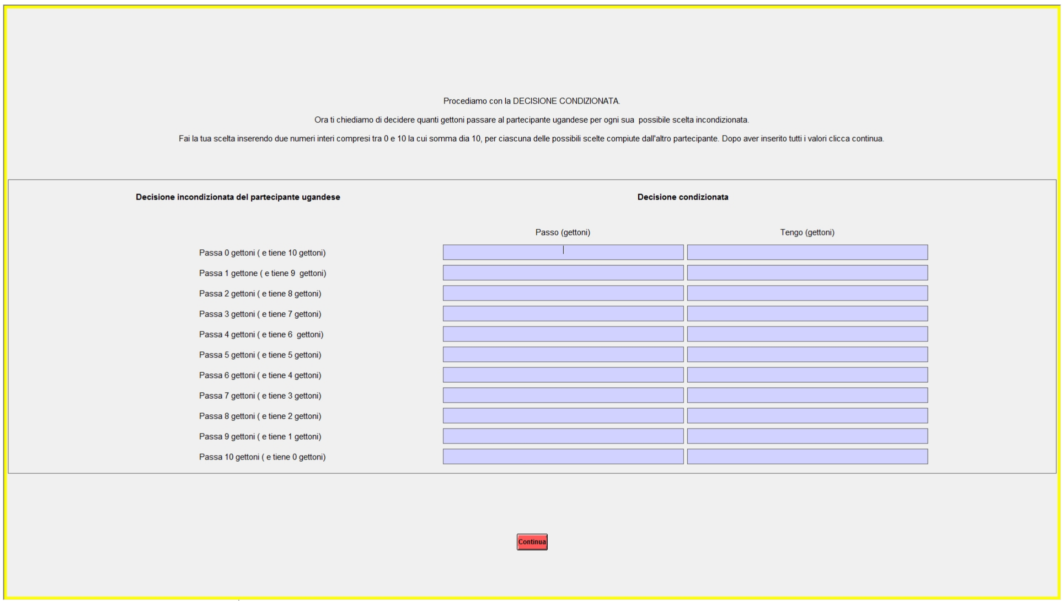

Figure C.13: Decision Screen: Conditional Choice

C.5 Instructions for Subjects in Uganda

INSTRUCTIONS FOR UGANDA SESSIONS (August 9 version).

Good morning and thank you for participating in our experiment.

You will receive 3000 UGX as show-up fee.

You will have to make a single, very simple decision. Depending on your decision and the decision of another person, you may receive an additional variable sum of money.

We will not deceive you. The decision you will make will not be revealed to anyone. All the information we shall give to you during the experiments is true.

At the beginning of the experiment you have been given a card with a number.
Please keep this card until the end of the experiment.

You will be matched with another person who lives in Trento, Italy, Europe who has been given a card with the same number of yours and who has made decision which is very similar to the one you are going to make.

Your decision:
A total of 10 tokens will be given to you. We ask you to decide how many tokens to keep for yourself and how many to give to this other Italian person with whom you have been paired. You will be asked to put the tokens you want to keep in an envelope with the label “TAKE” and the tokens you want to give to the other person in another envelop labelled “GIVE”.

You may decide for example to send 3 tokens to the other person. Alternatively, you may decide to keep all the 10 tokens for yourself, send them all to the other person or any other division you like.

At this point the game ends.

The person in Italy with whom you have been matched has already made a very similar decision. This person has already told us what he/she would do if he/she received the tokens. To this person we also asked what he/she would do if he/she where to know that you would pass him nothing, or one token, or two tokens and so on.

Only one choice - either yours or your Italian partner’s choice - will be selected as the actual choice that determines your payment.

Once you have made your choice, we will draw a piece of paper from a box containing five pieces with the word “Italy” and five with the word “Uganda”.

If we draw a piece of paper with “Uganda“ written on it, than your choice will be taken into account, otherwise we will consider the Italian participants’ choices. In the latter case, we will show you what your Italian partner has decided and we will pay you accordingly.

[PROBLEM: the best solution would be to make a single draw for all the Ugandan subjects, and not one draw for each session, is it feasible?]

For each token you earn, you will receive 700 UGX while your Italian partner will receive one euro for each token s/he earns. [more on euro/UGX exchange rate?]

Let’s make some examples.

Example 1.

You decide to take 8 tokens for yourself and to send 2 tokens to the Italian person you have been paired with. This is your choice. Your Italian partner has decided to take 6 tokens for himself and to send you the remaining 4 tokens. This is the Italian person choice.
After your decision we proceed with the draw to decide which choice to consider and we pick out a piece of paper with “Uganda” written on it. This means that your choice will be taken into account.
Your payment in this case will be of 8 tokens x 700 UGX=5600 UGX, while your Italian partner will be paid 2 tokens x 1 euro= 2 euro.

Example 2.

Suppose your choice and your Italian partner’s choices are the same as in the previous example, but now we pick out a piece of paper with “Italy” written on it. In this case we will take into account the Italian participant’s choice.

Since he decided to send you 4 tokens, your payment will be of 4 tokens x 700 UGX=2800 UGX, while his payment will be of 6 tokens*1 euro= 6 euro.

It is important that you understand that nobody will observe the decision you make.
You will need your number to collect your payment. One of us will record how much money you receive from the game according to this number (*point at "counter").* He will be the only one who knows how many tokens you put in the envelope but he won't know whose number goes with whom! He'll count out your payment. A different person will give you an envelope with the tokens you receive from the game according to this number. This person who sees your face and knows your number won't know what you decided. So your decision is totally anonymous.

We ask you to make your decision now
You will collect your payment right after the draw.

Bibliography

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

Andreoni, J. (1995). Cooperation in public-goods experiments: Kindness or confusion? *The American Economic Review*, 891–904.

Andreoni, J. & Bernheim, B. D. (2007). Social image and the 50-50 norm. *Econometrica*.

Andreoni, J. & Croson, R. (2008). Partners versus strangers: Random rematching in public goods experiments. In C. R. Plott & V. L. Smith (Eds.), *Handbook of Experimental Economics Results* (Vol. 1, pp. 776–783). Elsevier.

Andreoni, J., Miller, J. H. et al. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *Economic journal*, *103*(418), 570–585.

Andreoni, J. & Samuelson, L. (2006). Building Rational Cooperation. *Journal of Economic Theory*, *127*(1), 117–154.

Arifovic, J. & Ledyard, J. (2012). Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism. *Journal of Public Economics*, *96*(9), 808–823.

Artinger, F., Exadaktylos, F., Koppel, H., & Sääksvuori, L. (2010). *Applying quadratic scoring rule transparently in multiple choice settings: A note.* Jena Economic Research Papers.

Axelrod, R. (2009). *The evolution of cooperation: Revised edition.* Basic Books.

Axelrod, R. & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396.

Axelrod, R. (1980a). Effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, *24*(1), 3–25.

Axelrod, R. (1980b). More effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, *24*(3), 379–403.

Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics*, *11*(2), 122–133.

Bardsley, N. & Sausgruber, R. (2005). Conformity and Reciprocity in Public Good Provision. *Journal of Economic Psychology*, *26*(5), 664–681.

Ben-Ner, A., Putterman, L., Kong, F., & Magan, D. (2004). Reciprocity in a two-part dictator game. *Journal of Economic Behavior & Organization*, *53*(3), 333–352.

Bendor, J. & Swistak, P. (1995). Types of evolutionary stability and the problem of cooperation. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(8), 3596–3600.

Benz, M. & Meier, S. (2008). Do people behave in experiments as in the field?—evidence from donations. *Experimental Economics*, *11*(3), 268–281.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142.

Bolton, G. E. & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *The American Economic Review*, *90*(1), 166–193.

Bowles, S. & Gintis, H. (2011). *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton University Press.

Brandts, J. & Charness, G. (2011). The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics*, *14*(3), 375–398.

Brier, G. W. (1950a). Verification of forecasts expressed in terms of probability. *Monthey Weather Review*, *78*(1), 1–3.

Brier, G. W. (1950b). Verification of forecasts expressed in terms of probability. *Monthey Weather Review*, *78*(1), 1–3.

Burks, S. V., Carpenter, J. P., & Verhoogen, E. (2003). Playing both roles in the trust game. *Journal of Economic Behavior & Organization*, *51*(2), 195–216.

Burton-Chellew, M. N., El Mouden, C., & West, S. A. (2016). Conditional Cooperation and Confusion in Public-Goods Experiments. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(5), 1291–1296.

Burton-Chellew, M. N., Nax, H. H., & West, S. A. (2015). Payoff-based learning explains the decline in cooperation in public goods games. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1801), 20142678.

Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Çelen, B., Schotter, A., & Blanco, M. (2017). On blame and reciprocity: Theory and experiments. *Journal of economic theory*, *169*, 62–92.

Charness, G. (2004). Attribution and reciprocity in an experimental labor market. *Journal of labor economics*, *22*(3), 665–688.

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

Chaudhuri, A. (2011). Sustaining Cooperation in Laboratory Public Goods Experiments: a Selective Survey of the Literature. *Experimental Economics*, *14*(1), 47–83.

Clark, K. & Sefton, M. (2001a). The sequential prisoner's dilemma: Evidence on reciprocation. *The Economic Journal*, *111*(468), 51–68.

Clark, K. & Sefton, M. (2001b). The sequential prisoner's dilemma: Evidence on reciprocation. *The Economic Journal*, *111*(468), 51–68.

Cooper, D. J. & Kagel, J. (2011). Other regarding preferences: A survey of experimental results. In J. Kagel & A. Roth (Eds.), *The Handbook of Experimental Economics* (Vol. 2). Princeton: Princeton University Press.

Cooper, D. J. & Stockman, C. K. (2002). Fairness and learning: An experimental examination. *Games and Economic Behavior*, *41*(1), 26–45.

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

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, *100*(2), 193–201.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80.

Diekmann, A. (2004). The power of reciprocity: Fairness, reciprocity, and stakes in variants of the dictator game. *The Journal of Conflict Resolution*, *48*(4), 487–505.

Doebeli, M. & Knowlton, N. (1998). The evolution of interspecific mutualisms. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(15), 8676–8680.

Eckel, C. C. & Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and Economic Behavior*, *16*(2), 181–191.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583–610.

Engelmann, D. & Fischbacher, U. (2009). Indirect reciprocity and strategic reputation building in an experimental helping game. *Games and Economic Behavior*, *67*(2), 399–407.

Engle-Warnick, J. & Slonim, R. L. (2006). Learning to trust in indefinitely repeated games. *Games and Economic Behavior*, *54*(1), 95–114.

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

Fallucchi, F., Luccasen, R. A., & Turocy, T. L. (2018). Identifying discrete behavioural types: A re-analysis of public goods game contributions by hierarchical clustering. *Journal of the Economic Science Association*, 1–17.

Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? an experimental investigation. *The Quarterly Journal of Economics*, *108*(2), 437–459.

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

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

Fischbacher, U. & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, *100*(1), 541–56.

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

Fischbacher, U., Gächter, S., & Quercia, S. (2012). The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology*, *33*(4), 897–913.

Fisman, R., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *The American Economic Review*, *97*(5), 1858–1876.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*(3), 347–369.

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *The Review of Economic Studies*, *38*(1), 1–12.

Fudenberg, D. & Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, *54*(3), 533–534.

Fundenberg, D. & Maskin, E. (1990). Evolution and cooperation in noisy repeated games. *The American Economic Review*, *80*(2), 274–279.

Gächter, S. & Falk, A. (2002). Reputation and reciprocity: Consequences for the labour relation. *The Scandinavian Journal of Economics*, *104*(1), 1–26.

Gächter, S. & Herrmann, B. (2009). Reciprocity, culture and human cooperation: Previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1518), 791–806.

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

García, J. & van Veelen, M. (2016). In and out of equilibrium i: Evolution of strategies in repeated games with discounting. *Journal of Economic Theory*, *161*, 161–189.

García, J. & van Veelen, M. (2018). No strategy can win in the repeated prisoner's dilemma: Linking game theory and computer simulations. *Frontiers in Robotics and AI*, *5*, 102.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. i. *Journal of Theoretical Biology*, *7*(1), 1–16.

Hartig, B., Irlenbusch, B., & Kölle, F. (2015). Conditioning on what? heterogeneous contributions and conditional cooperation. *Journal of Behavioral and Experimental Economics*, *55*, 48–64.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., … et al. (2005). "economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, *28*(6), 795–815.

Herrmann, B. & Thöni, C. (2009). Measuring Conditional Cooperation: a Replication Study in Russia. *Experimental Economics*, *12*(1), 87–92.

Hirshleifer, J. & Coll, J. C. M. (1988). What strategies can support the evolutionary emergence of cooperation? *The Journal of Conflict Resolution*, *32*(2), 367–398.

Hoffman, E., McCabe, K., & Smith, V. L. (1996). Social distance and Other-Regarding behavior in dictator games. *The American Economic Review*, *86*(3), 653–660.

Houser, D. & Kurzban, R. (2002). Revisiting Kindness and Confusion in Public Goods Experiments. *American Economic Review*, *92*(4), 1062–1069.

Imhof, L. A., Fudenberg, D., & Nowak, M. A. (2005). Evolutionary cycles of cooperation and defection. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(31), 10797–10800.

Iriberri, N. & Rey-Biel, P. (2011). The role of role uncertainty in modified dictator games. *Experimental Economics*, *14*(2), 160–180.

Isaac, R. M., McCue, K. F., & Plott, C. R. (1985). Public goods provision in an experimental environment. *Journal of Public Economics*, *26*(1), 51–74.

Kelley, H. H. & Stahelski, A. J. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology*, *16*(1), 66.

Kim, O. & Walker, M. (1984). The free rider problem: Experimental evidence. *Public Choice*, *43*(1), 3–24.

Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J., & Sutter, M. (2008). Conditional Cooperation on Three Continents. *Economics Letters*, *101*(3), 175–178.

Kropotkin, P. (1902). *Mutual aid: A Factor of Evolution*. McClure Phillips & Co, New York.

Krupka, E. & Weber, R. A. (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic Psychology*, *30*(3), 307–320.

Kurzban, R. & Houser, D. (2005). Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences*, *102*(5), 1803–1807.

Lazear, E. P., Malmendier, U., & Weber, R. A. (2012). Sorting in experiments with application to social preferences.

Ledyard, J. O. (1994). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *Handbook of Experimental Economics* (pp. 112–193). Princeton University Press.

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

List, J. A. (2007). On the interpretation of giving in dictator games. *The Journal of Political Economy*, *115*(3), 482–493.

Neugebauer, T., Perote, J., Schmidt, U., & Loos, M. (2009). Selfish-Biased Conditional Cooperation: On the Decline of Contributions in Repeated Public Goods Experiments. *Journal of Economic Psychology*, *30*(1), 52–60.

Nishi, A., Christakis, N. A., & Rand, D. G. (2017). Cooperation, decision time, and culture: Online experiments with american and indian participants. *PloS one*, *12*(2), 1–9.

Nowak, M. & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, *364*(6432), 56–58.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*(5805), 1560–1563.

Nowak, M. & Highfield, R. (2011). *Supercooperators: Altruism, evolution, and why we need each other to succeed*. Simon and Schuster.

Nowak, M. & Sigmund, K. (1989). Oscillations in the evolution of reciprocity.

Palfrey, T. R. & Prisbrey, J. E. (1997). Anomalous behavior in public goods experiments: How much and why? *The American Economic Review*, *87*(5), 829–846.

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

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*, 427–430.

Roberts, G. & Sherratt, T. N. (1998). Development of cooperative relationships through increasing investment. *Nature*, *394*(6689), 175–179.

Seabright, P. (2010). *The company of strangers: A natural history of economic life - revised edition*. Princeton University Press.

Selten, R. & Hammerstein, P. (1984). Gaps in harley's argument on evolutionarily stable learning rules and in the logic of "tit for tat". *The Behavioral and Brain Sciences*, *7*(1), 115–116.

Selten, R. & Stoecker, R. (1986). End behavior in sequences of finite prisoner's dilemma supergames a learning theory approach. *Journal of Economic Behavior & Organization*, *7*(1), 47–70.

Sen, S. (1996). Reciprocity: A foundational principle for promoting cooperative behavior among self-interested agents. In *Proceedings of the Second International Conference on Multiagent systems* (Vol. 315321).

Servátka, M. (2009). Separating reputation, social influence, and identification effects in a dictator game. *European Economic Review*, *53*(2), 197–209.

Sherratt, T. N. & Roberts, G. (2002). The stability of cooperation involving variable investment. *Journal of Theoretical Biology*, *215*(1), 47–56.

Sugden, R. (1984). Reciprocity: The supply of public goods through voluntary contributions. *The Economic Journal*, *94*(376), 772–787.

Szolnoki, A. & Perc, M. (2012). Conditional strategies and the evolution of cooperation in spatial public goods games. *Physical Review E*, *85*(2 Pt 2), 026104.

Tinghog, G., Andersson, D., Bonn, C., Bottiger, H., Josephson, C., Lundgren, G., … Johannesson, M. (2013). Intuition and cooperation reconsidered. *Nature*, *498*, E-1–E-3.

Tonin, M. & Vlassopoulos, M. (2013). Experimental evidence of self-image concerns as motivation for giving. *Journal of economic behavior & organization*, *90*, 19–27.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.

van Veelen, M., García, J., Rand, D. G., & Nowak, M. A. (2012). Direct reciprocity in structured populations. *Proceedings of the National*

Academy of Sciences of the United States of America, *109*(25), 9929–9934.

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

Vu, T. T. T. (2018). The wolf in sheep's clothing - an experimental analysis on signalling social preferences.

Wahl, L. M. & Nowak, M. A. (1999a). The continuous prisoner s dilemma: I. linear reactive strategies. *Journal of Theoretical Biology*, *200*, 307.

Wahl, L. M. & Nowak, M. A. (1999b). The continuous prisoner s dilemma: II. linear reactive strategies with noise. *Journal of Theoretical Biology*, *200*, 323.

Wilson, E. O. (2012). *The social conquest of earth*. WW Norton & Company.

Zhang, H. & Perc, M. (2016). Evolution of conditional cooperation under multilevel selection. *Scientific Reports*, *6*, 23006.