Università Commerciale "Luigi Bocconi"
PhD School

PhD program in Economics & Finance
Cycle: 32^{nd}
Disciplinary Field (code): SECS-P/01

Essays in Experimental and Behavioral Economics

Advisor: Pierpaolo Battigalli
Co-Advisor: Giuseppe Attanasi

PhD Thesis by
Alessandro Stringhi
ID number: 3028132

Year: 2022

Acknowledgments

I want to take a brief moment to thank all the professors and fellow students that helped me during this journey towards the completion of this Ph.D thesis.

Foremost, I would like to express my sincere gratitude to my advisor, Prof. *Pierpaolo Battigalli*, for his guidance and constant supervision as well as for all the knowledge that he shared with me.

With the hope of working with him again, I extend my gratitude also to my former advisor, Prof. *Paolo Pin*, which firstly introduced me to this field of studies.

This thesis would not have been possible without the help of my second advisor, Prof. *Giuseppe Attanasi*, which gave me valuable advice and financial support for running the experiments.

Alongside him, I want to thank the professors and fellow researchers that work at LEEN for the invaluable help I received during my stay in Nice, in particular, my co-author *Sara Gil-Gallen*.

I would also like to acknowledge the helpful criticism from the reviewers, Prof. *Pablo Brañas-Garza* and *Nikolaos Georgantzis*

I would like to acknowledge the financial and academic support of the Bocconi University and its staff, in particular I acknowledge the financial support from *Fondazione Invernizzi*, and the passion and time invested by the directors of the program, Prof. *Max Croc*e and *Marco Ottaviani*.

Finally, I thank my cohort for these years spent together. I wish them the most fulfilling and happy life.

Contents

Introduction

In this thesis I investigate human behavior in laboratory experiments. The aim of my investigation is to improve our understanding on preferences and behavior. My work is comprised of two separate studies. The first study investigates preferences, with special attention to preferences for equity or guilt avoidance, in the context of a Trust Game. In this study I propose a design able to test two theories one against the other, and all my predictions are justified by a rigorous theoretical analysis. In this study I further investigate the findings of Charness & Dufewnberg (2006), a paper in which the authors presented evidences in favor of the model of guilt aversion, but their interpretation was lately challenged. This study contributes to the theoretical literature on psychological games and on exotic preferences. I show, with a theoretical analysis, that guilt and inequity aversion models behave in opposite ways when the game's payoffs are manipulated. Indeed, I designed an experiment to test the two models that relies on a simple payoff manipulation. The second study proposes a framing able to increase cooperation in a Prisoner's Dilemma. The framing proposed is the one of a Tournament. In this tournament a pair of subjects may coordinate in order to defeat the opposing pair. This framing aligns the incentives of the two players involved, since they became allies fighting a common enemy. This framing prove itself very successful in coordinating the two players and it leads to a higher cooperation, mitigating the undesirable situation associated with games like the Prisoner's Dilemma. The findings of this study are meaningful because the competitive framing can be easily implemented in various real life situations, and possibly, it can deliver desirable results.

Chapter 1

Guilt and Fairness

Guilt and Fairness

Alessandro Stringhi*[1]

[1]Università Bocconi, Italy

Abstract

Our society is built upon trust, but the reasons why we should trust a person are not always transparent. To answer to this question we need to first understand what drives someone to repay the trust that we put in him. Many theories try to explain those motivations, with inequity and guilt aversion among the most prominent. With this work I aim to identify the main driver of trustworthiness. Building on Charness & Dufwenberg (2006) framework I design an experiment that allows me to test separate theories against one another. I achieve this goal by increasing the material payoffs of the trustor in the treatment. My behavioral predictions across treatments are derived from an analysis of the theoretical models under investigation without relying on equilibrium analysis.

1 Introduction

Personal and economic relationships are built on trust. Whom should I trust? Why should I trust someone? How can I prove myself trustworthy? All these questions are of great importance. Charness & Dufwenberg *Econometrica* (2006), C&D from now on, dig into the problem, showing with an experiment how trust, communication and guilt are interconnected. The authors used a trust game with hidden actions to ascertain whether communication is able to increase cooperation. Their explanation is that a promise made by the trustee increases the expectation of the trustor about her material payoff. If the trustee is guilt averse he will feel more guilty by taking all the profit after sending a promise, increasing the likelihood that he will keep

*alessandro.stringhi@phd.unibocconi.it

his promise. The authors were able to confirm their predictions, in the communication treatment there was a higher frequency of cooperative actions played by both the trustee and the trustor. They suggested that a model of guilt aversion well explains the evidences that communication improves trust, cooperation and efficiency.

Despite the success of this paper the authors weren't able to provide strong evidences in favor of guilt aversion and the necessity to include beliefs into the utility function of a decision maker. They acknowledged some other critiques in a follow up paper (Charness & Dufwenberg (2010)). Vanberg (2008) challenged the findings of the first paper suggesting that "the effects of promises cannot be accounted for by changes in payoff expectations. This suggests that people have a preference for promise-keeping per se". Other authors also suggest that the cost of lying is belief independent (Ellingsen & Johannesson (2004), Chen et al. (2008), Kartik (2009)) and it relies on the inequity aversion of the trustor.

I aim to dig deeper into the subject, trying to answer the underlying question that the aforementioned papers did not explicitly addressed, namely which model is better suited to explain the behavior of people in a trust game. To do so I propose an experiment built on C&D's framework. With this experiment I expect to test simple guilt (Battigalli and Dufwenberg (2007)) versus inequity aversion (Fehr and Schmidt's (1999)). In my experiment I propose a version of the Trust Minigame used by C&D, but without hidden actions. This choice allows me to test the model of simple guilt without having to worry about other confounding factors such as shame or image concerns, since Tadelis (2007) already showed the importance of their role in C&D game. The mechanism that allows me to disentangle the two theories is a payoff manipulation in the treatment group, namely increasing the payoff of the trustor in every terminal history. In the control group, as in C&D (2006), both players earn the same amount in every terminal history that does not end with the trustee grabbing all the surplus. In the treatment, instead, the trustor's earnings are higher than those in the control, while the ones of the trustee are left unchanged. Therefore in the treatment there is always inequality between the monetary payoffs of the trustee and the trustor. With this simple manipulation it is possible to compare the predictions of the inequity aversion model versus those of the guilt aversion model. The former model postulates that, when the profit is shared, only in the treatment the trustee suffers a disutility caused by his inferiority aversion, since now the trustor earns more than him when the profit is shared. This will lead to a decrease in the frequency of cooperation. The guilt aversion model predicts the opposite result. Since the material payoff of the trustor is higher in the treatment, also it is the value that she expects to receive. The higher expectation of the trustor translates in a higher guilt for the trustee if he betrays the trust put in him by his partner.

I justify my predictions by analyzing the two models separately. As solution concept I rely on strong rationalizability and I assume incomplete information about the personal traits of the players. This choice better reflects the reality of a controlled experiment. Moreover I cannot rely on equilibrium concepts that postulates correct beliefs of the players because player's beliefs about the true preferences are themselves under investigation. In each separate analysis I assume that players have strong beliefs about only one preference at the time. This assumption facilitates my analysis and highlights the fundamental differences of the two models. Therefore, since it is completely unrealistic to assume that those beliefs are correct, the use of equilibrium concepts is not justified in this framework.

This paper is organized as follow. In section 2 I discuss the relevant literature. In section 3 I analyze the two models and I prove that the two models behave differently when the payoffs of the trustor are increased. Section 4 presents the experimental design and in Section 5 I present the behavioral predictions derived from the theoretical analysis. In section 6 I show the results of the experiment. Section 7 concludes.

2 Related Literature

This paper draws inspiration from the experiment proposed in Charness & Dufwenberg (2006) and from the discussion generated by it. In C&D (2006) the authors designed an experiment in order to study the role of pre-play communication in a Trust Minigame. According to standard Game Theory, non-binding pre-play communication shouldn't have any effect on the behavior of players in a Trust Game. The authors theorized that, if the motivations of the players are affected by a disposition to experience guilt when letting down the expectations of others, then communication should play an important role. The authors argued that a message from the trustee, in which he makes a promise to share the profit, should increase the expectation of the trustor. If the trustee is averse to guilt, then his utility depends also on the beliefs of the trustor. Since the promise increased the expectation of the trustor, now the trustee is more inclined to keep his promise of sharing, and the trustor now has a reason to trust him and to invest. To test this hypothesis the authors designed an experiment in which participants play a Trust Minigame with hidden actions. The trustor could play *Out*, ending the game by giving 5$ to both players, or play *In* and let the trustee make his choice. If the trustor played *In*, the trustee had to choose to *Roll*[1], and equally share a profit of 20$, or *Don't Roll* and grab the surplus, 14$, leaving the trustor with nothing.

[Figure 1 here]

The main treatment of the experiment involved a pre-play communication phase, in which the trustee had the opportunity to send a message to the trustor. Usually this opportunity was used to send promises to play *Roll*. The authors observed a significant increase in the frequency of *Roll* and *In* in treatments with communication. Also the measured first-order beliefs of the trustor and second-order beliefs of the trustee were significantly higher in the communication treatments. Moreover they found a correlation between beliefs and behavior, particularly for the trustees. The trustees who choose *Roll* had a significantly higher measured second-order beliefs. All the experimental evidences are in agreement with all the predictions given by the guilt aversion model, as pointed out by C&D. Despite that, the authors do not rule out alternative explanations for their findings.

[1] In the actual experiment a die was rolled. With 5/6 probability the trustor earns 12$ and with a 1/6 probability she earns 0$. I disregard the implications of die roll in this discussion because I removed the chance move in my experiment. Ignoring the chance move doesn't undermine the validity of C&D's results.

This is pointed out in Vanberg (2008), where the author argues that the differences between treatments may reflect a preference for promise keeping. Moreover, the correlation between trustee's behavior and second-order beliefs can be caused by a false consensus effect, namely subjects believe that others can well predict their behavior. Vanberg claims that C&D's experiment fails to distinguish between what he refers as *commitment-based* and *expectation-based* explanation for promise keeping. The latter is the explanation given by C&D, in which promises are kept because they increase expectation about future payoffs, leading to more trustworthiness. While the former claims that players are not only concerned about the expected consequences of their behavior, but also about the fulfillment of obligations based on previously set agreements. This notion that promises induce emotional commitments to fulfill obligations has been adopted by many authors (Braver (1995), Ostrom, Walker, and Gardner (1992), Ellingsen and Johannesson (2004)). In particular Ellingsen and Johannesson (2004) propose a model of social preferences that includes a "taste (...) for keeping one's word". The model assumes a fixed cost of lying, consistent with the *commitment-based* explanation, and it relies on inequity aversion. This means that there are two possible models that can explain the effect of communication on trustworthiness, but it is hard to suggest one model over the other if it is not possible to distinguish between the two explanations for promise keeping. In order to investigate this issue, Vanberg (2008) designed an experiment which was able to distinguish between *commitment* and *expectation-based* explanation. Vanberg's experiment allows for an independent variation in promises and beliefs, this is achieved by randomly rematching part of the subjects after the pre-play communication phase. In the experiment the subjects play a Dictator Minigame, where the choice of the dictator is the same choice that faces the second mover in C&D's experiment. At the beginning of each round, each subject is paired with another and they have the opportunity to communicate and exchange promises in case they will be choose as dictator. After the communication phase each subject is randomly assigned to his role, dictator or recipient, and then the dictator makes his choice. In the main treatment of the experiment some dictators are switched, therefore they end up playing with a recipient with whom they haven't talk previously. The dictators are informed of the switch, while the recipients are not. Moreover the dictators have the opportunity to read the messages received by the new recipient, message which was sent by a different dictator. The switch allows to make predictions on the motivation of the dictators to keep promises. According to the *expectation-based* explanation the behavior of the dictator in the switch and non-switch situations should be identical, since his guilt depends only on the recipient's expectation, which is influenced by the promises that he received, regardless who made it. While the *commitment-based* explanation for promise keeping suggests that a dictator's behavior should be affected only by his own promises and only in the non-switch condition. The results of the experiment showed that second-order beliefs did not differ significantly between the switch and non-switch condition, but dictators were significantly more likely to play *Roll* in the non-switch condition. These results are inconsistent with the thesis of C&D (2006) and they instead suggest that dictators are motivated primarily by their own promises. Although the results show evidences in favor of the *commitment-based* explanation for promise keeping, Vanberg does not discard the guilt aversion model. The goal of the experiment was to distinguish between and to test two explanations for promise keeping, not the models implied by those explanations. More evidences in favor of the *commitment-based* explanation were found by Di Bartolomeo et al. (2019) in a refined version of Vanberg's experiment. Despite that, Di Bartolomeo et al. (2020) founds evidence in favor of *expectation-based* explanation in a new experiment in which they implemented that partner-switch mechanism in C&D design.

Although we are still in uncharted territory, in my paper I do not investigate the role of communication, but I focus on the implicit question never fully addressed by previous works. Does guilt aversion performs better than inequity aversion in explaining the experimental evidences? Is guilt aversion, and therefore psychological game theory, necessary? Or it is possible to explain the same phenomena with a more standard approach? My research question contributes to the literature by tackling directly the issue. Previous papers used an indirect approach, they tried to find evidences compatible with a model, but without trying to falsify the other. Understanding which model is more suited to model a Trust Game is necessary for both theory and applications.

3 Theory

In this section I make a comparison between two prominent models, the Fehr & Schmidt inequity aversion and Battigalli & Dufwenberg guilt aversion. Although this two models are able to predict a positive fraction of prosocial outcomes in a Trust Minigame, their predictions vary when the payoffs of the first mover are manipulated. I show that when the payoff of the trustor is higher than the payoff of the trustee, the inequity aversion model predicts a low frequency of cooperative outcomes. Instead, the guilt aversion model will predict an higher frequency and the difference between the two models become starker as the trustor's payoffs increase.

I use these theoretical predictions to justify my experimental design in order to understand which model is better suited to explain people's behavior. I start by introducing a parametrized version of the Trust Minigame, I use this parametrization to show the different monotonicity properties of the two models. Then I briefly explain the solution concept that I use and then I define all the relevant mathematical objects necessary in the analysis. At last I analyze the two models separately.

3.1 Trust Minigame

The Trust Minigame is a dynamic game that models the behavior of two persons during an investment decision. In this game, the trustor has to choose if to invest or not in the trustee's project. If the investment is made, the trustee can keep all the profit for himself, or share the profit with the trustor. In the game commonly refer as Trust Game[2] the trustor can choose to invest any fraction of her endowment, thus is a compact continuous game. Here I consider a simpler version[3] in which the decision to invest is binary, therefore is referred to as Minigame. Here I introduce a parametrized version of the Trust Minigame. The a parameter $m \in \mathbb{R}$, with $m \geq 1$, rescales the trustor's material payoffs and represent the ratio between trustor and trustee's payoffs in all terminal histories, except when the trustee takes all the profit.

The game tree of the parametrized Trust Minigame is depicted in Figure 2.

[Figure 2 here]

Before defining the game form I have to make a remark. The objective of this section is to compare the predictions of two different models. Both these

[2] The game was originally called the Investment Game by Berg et al. (1995).

[3] This earlier and simpler game was named Trust Game by Kreps (1990).

models postulate that players have personal traits that affect their behavior. Assuming that players know the personal traits of the other players participating in the experiment is completely unrealistic, therefore I define the game as a multistage game with payoff uncertainty and observable actions.

The Trust Minigame is a two players game, thus $I = \{A, B\}$, where A is the trustor and B is the trustee. Let be $\overline{H}$ the set of feasible histories, partitioned in two sets, $H = \{\varnothing, In\}$ and $Z = \{Out, (In, Sh), (In, Tk)\}$, which are the set of non-terminal and terminal histories respectively. From now on I write Sh and Tk instead of (In, Sh) and (In, Tk) for convenience. Player A is the first mover and she is active at the root, therefore her set of feasible actions is $\mathcal{A}_A(\varnothing) = \{In, Out\}$. While player B is active only if player A plays In, therefore $\mathcal{A}_B(In) = \{Sh, Tk\}$.

Let $\Theta = \times_{i \in I} \Theta_i$ be the space of personal traits. Since in this paper I study both guilt and inequity aversion models, Θ contains the parameters for both models, therefore $\Theta_i = \Theta_i^G \times \Theta_i^S \times \Theta_i^I$, where Θ^G is the space of guilty sensitivity parameters and Θ^S and Θ^I are the spaces of superiority and inferiority sensitivity parameters. Later in this section I will define the structure of Θ, making assumptions on the values of the parameters according to each model under investigation.

Let $\Delta_2 = \Pi_{j \in I} \Delta_{j,2}$ be the space of second-order conditional probability systems (CPS). In order to define this space I follow the procedure of Corrao, Battigalli, & Dufwenberg (2019), from now on CBD, and I defer to that paper for a more general definition. To define Δ_2 I need to make a step back and define the space of first-order CPSs Δ_1. Each player i is uncertain about the outcome of the game and the coplayer's personal traits, therefore the primitive uncertainty space is $Z \times \Theta_{-i}$. Thus a first-order belief of i is an element of $\Delta(Z \times \Theta_{-i})$. The strategic reasoning of each player is represented using their beliefs conditional on each personal history[4]. Therefore I use the notation $\alpha_i(Sh|In)$ to denote the probability that player i assigns to *Share* given that history *In* is reached, and $\alpha_i(\theta_{-i} > \theta|In)$ to denote the probability that i assigns to $-i$'s sensitivity to be greater than a certain value given that history *In* is reached. Now consider the set of all maps from H to the belief set $\Delta(Z \times \Theta_{-i})$, which is $[\Delta(Z \times \Theta_{-i})]^H$. I say that $\alpha_i \in [\Delta(Z \times \Theta_{-i})]^H$ is a first-order conditional probability system (CPS) if it satisfies the following properties:

1 **Knowledge implies Belief**: for every $h \in H$, $\alpha_i(h|h) = 1$;

[4]Given the simple structure of this game the set of personal histories coincide with the set of all histories.

2 **Chain rule**: for all $h,\ h' \in H$ and $F \subseteq Z(h') \times \Theta_{-i}$,

$$h \preceq h' \implies \alpha_i(F|h) = \alpha_i(F|h')\alpha_i(h'|h),$$

where $Z(h)$ is the set of terminal histories consistent with history h;

3 **Own-action independence (OAI)**: for all $h \in H$, $a_i, a_i' \in \mathcal{A}_i(h)$, $a_{-i} \in \mathcal{A}_{-i}(h)$ and $G \subseteq \Theta_{-i}$,

$$\alpha_i(Z(h, \alpha_{-i}) \times G|h, a_i) = \alpha_i(Z(h, \alpha_{-i}) \times G|h, a_i').$$

I call $\Delta_{i,1} \subseteq [\Delta(Z \times \Theta_{-i})]^H$ the space of first-order CPSs and $\Delta_1 = \Pi_{i \in I}\Delta_{i,1}$. Using a triple $(z, \theta, \alpha) \in Z \times \Theta \times \Delta_1$ is possible to describe the behavior of a player. Since players are uncertain about the behavior of other players, i forms beliefs over $Z \times \Theta_{-i} \times \Delta_{-i,1}$. I let $\beta_i(\cdot|h) \in \Delta(Z \times \Theta_{-i} \times \Delta_{-i,1})$ denote a generic second-order belief of i conditional to history h. In a similar fashion I define $\Delta_{i,2}$ as the space of second-order CPSs as the space of second-order beliefs that satisfy properties 1-2-3. As in BCD I do not use higher order beliefs, because the utility functions of the models under investigation depend only on first-order beliefs. It is important to note that it is enough to define only the second-order CPSs because the first-order CPSs can be recovered by marginalization.

Let Y be the set of material outcomes, I define the outcome function $\pi^m : Z \to Y$. This outcome function is parametrized by m and each value of m defines a different game form. The outcome function of player B is independent from m, while the outcome function of player A is scaled as shown in Figure 2.

Here I do not define the players' utility functions u_i in details because they are model specific and they will be defined separately later. In principles, the utility function depends on the outcome, personal traits, and possibly, beliefs, thus $u_i : Z \times \Theta \times \Delta_2 \to \mathbb{R}$.

Having defined all the ingredients I use the following mathematical structure to define the parametrized Trust Minigame

$$\Gamma_m = \left\langle I, \Theta_0, (\Theta_i, A_i, \mathcal{A}_i(\cdot), \Delta_{i,2}, \pi_i^m, u_i)_{i \in I} \right\rangle. \tag{1}$$

Note that the subscript m defines a different game for each value of $m \geq 1$ through the parametrized outcome function $\pi_A(m)$. Having defined the game I proceed defining the solution concept that I will use in my analysis.

3.2 Strong Beliefs and Strong Rationalizability

The solution concept that I use in my analysis is strong rationalizability, using an approach similar to BCD. Strong rationalizability (cf. Battigalli and Siniscalchi 2002) is a solution concept consistent with forward induction reasoning and it is based on the notion of strong beliefs. I say that player i strongly believes that an event $F \neq \varnothing$ is true if and only if she is certain of F at all histories consistent with F, and I use the notation $SB_i(F)$ to express that. I also use the notation $SB(F)$ to say that every player strongly belief F. Moreover I adopt the use of an auxiliary operator, $CSB(F)$, which is defined as

$$CSB(F) = F \cap SB(F),$$

which means that everybody strongly belief F and such belief happens to be correct. I use the n^{th} order correct strong belief operator, $CSB^n(R)$, to describe the strategic sophistication of a player. $CSB^0(R) = R$ means that players are simply rational. $CSB^1(R) = R \cap SB(R)$ means that players are rational and strongly belief that everybody is rational. At last, $CSB^2(R) = R \cap SB(R) \cap SB(CSB^1(R))$ means that everybody is rational, everybody beliefs that everybody is rational and everybody beliefs that everybody beliefs that everybody is rational until this hypothesis is contradicted. For a more formal definition of all the operators used I suggest to refer to Battigalli & Siniscalchi (2002).

During the whole analysis I assume that the players have the highest level of sophistication necessary, namely $CSB^2(R)$. The procedure that I use is an iterated elimination of utility-relevant states. The procedure is defined as follow:

- (Step 0): For every $i \in I$, let $P_i^0 = Z \times \Theta_i \times \Delta_{i,1}$ and $P_{-i}^0 = Z \times \Theta_{-i} \times \Delta_{-i,1}$;
- (Step $n > 0$): For every $i \in I$, P_i^n is the set of triple $(z, \theta_i, \alpha_i) \in Z \times \Theta_i \times \Delta_{i,1}$ such that there exist a $\beta_i \in \Delta_{i,2}$ satisfying
  1 **Coherence:** $marg_i(\beta_i) = \alpha_i$;
  2 **Rationality:** Player i is rational at (z, θ_i, β_i);
  3 **Strong belief:** For every $\nu \in \{1, \ldots, n-1\}$, β_i strongly believes P_{-i}^{ν}.

A triple (z, θ_i, α_i) is said to be strong rationalizable if it survives all the iterations.

The primary focus of my analysis is to study the level of cooperation predicted by the two models in the control and in the treatment. Therefore I divide the set of strongly rationalizable triples into two subsets, C_i^n and S_i^n, which are the sets of cooperative and selfish triples. The set C_i^n is the set of all triples in which A plays *In* or B plays *Share* and the set S_i^n is the set of all triples in which A plays *Out* or B plays *Take*. Those two sets satisfy the following properties:

$$C_i^n \cap S_i^n = \emptyset, \quad C_i^n \cup S_i^n \subseteq P_i^n.$$

In the following analysis I will prove that these two sets have monotonicity properties with respect to the game parameter m. I will also prove that the two models predict opposite monotonicities. I will use the monotonicity properties of those sets to test the predictive power of the two models in the laboratory, since I set $m = 1$ in the control and $m = 1.5$ and $m = 2$ in the treatments. A larger set will translate in a higher frequency of participants playing the cooperative or selfish actions.

3.3 Inequity Aversion

3.3.1 The model

The model of inequity aversion was introduced in Fehr & Schmidt (1999). This model was designed to explain the evidence that fairness affects the behavior of people. Inequity aversion is defined as an other regarding preference because the material payoff of player j enters in the utility function of player i[5]. By inequity aversion the authors meant that in specific situations a person is willing to sacrifice part of his material payoff for a reduction in inequality between himself and his counterpart. More precisely they defined their preference as self-centered inequity aversion, meaning that only matters the inequality between the individual and the coplayer, and that the individual dislikes inequality more if it is disadvantageous for him.

Inequality can take two forms, either i earns more than j, or viceversa, leading to a advantageous and disadvantageous inequality for i respectively. Fehr-Schmidt model includes both instances of inequality. The authors formally defined self-centered inequality aversion as follows. Let be $I = \{i, j\}$ the set of players and $y_i = \pi_i(z)$ the material outcome associated with the terminal history $z \in Z$. Then the utility function for player i is given by

$$u_i(z) = y_i - \theta_i^I (y_j - y_i)^+ - \theta_i^S (y_i - y_j)^+, \tag{2}$$

where $(x)^+$ stand for the positive part of x and it is equivalent to $\max(x, 0)$.

The first term of the equation is simply the material payoff of player i. The second term of the equation captures the loss from the disadvantageous inequality suffered by i and θ_i^I is a parameter that measure the amount that

[5]To ease the exposition I will present the model for only two players.

i is willing to pay to reduce the inequality by 1. From now on I refer to θ^{I}[6] as the sensitivity parameter to inferiority aversion. This term of the equation is able to explain why some people is willing to reject a greedy, but still positive, offer in a ultimatum game.

The third term of (2) captures the loss from the advantageous inequality and θ^{S}[7] is the sensitivity parameter to superiority aversion. This term of the equation explains why some people offers a positive amount in a dictator game and I refer to this term as the altruistic component of the preference.

3.3.2 Beliefs and Type structure

In order to test the predictive power of the inequity aversion model, here I analyze the game using the inequity aversion model as working hypothesis. Namely, I assume that preferences for equity are the only relevant preferences and everybody agrees on that. Moreover I assume that player A is selfish, so it is more precise to refer to this model as role dependent inequity aversion, since this preference become relevant only if the agent play in the role of trustee. I call this event E, short for *Equity*. Formally this event is defined as

$$\begin{aligned} E &= \left(\Theta_A^G \times \Theta_A^S \times \Theta_A^I\right) \times \left(\Theta_B^G \times \Theta_B^S \times \Theta_B^I\right) \\ &= (\{0\} \times \{0\} \times \{0\}) \times \left(\{0\} \times [\underline{\theta}_B^S, \overline{\theta}_B^S] \times [\underline{\theta}_B^I, \overline{\theta}_B^I]\right) \end{aligned} \tag{3}$$

where $\underline{\theta}_B^S,\ \underline{\theta}_B^I \geq 0$ and $\overline{\theta}_B^S,\ \overline{\theta}_B^I \leq 1$. Since I conduct the analysis using role dependent inequity aversion as working hypothesis, this is equivalent to assume $CSB^2(E)$. In words, this means that B is inequity averse, A strongly believes that B is inequity averse and B strongly believes that A strongly believes that B is inequity averse. Same holds for the selfishness of A.

Having defined the model and the working hypothesis I proceed to analyze the game, showing that games with an higher m have a smaller set of cooperative players.

3.3.3 Analysis

Here I prove that the fraction of player that play the cooperative actions, *In* or *Share*, is smaller in those games with higher m. To do so, I first find the utility-relevant states consistent with strong rationalizability, then I divide

[6]In Fehr-Schmidt (1999) this parameter is called α. I will use this notation through the entirety of this paper in order to avoid confusion with other mathematical objects that I define using Greek letters.

[7]In Fehr-Schmidt (1999) this parameter is called β.

those states into two sets, the set of cooperative and selfish actions. At the end, I show that those sets have monotonicity properties with respect to m.

Strong Rationalizability The first three iterations of strong rationalizability give the key predictions:

1. The first iteration of the algorithm eliminates all the utility-relevant states that violates rationality. A's expected utility is

$$\alpha_A(In)\Big(\alpha_A(Sh)\cdot 10m\Big)+\alpha_A(Out)\Big(5m\Big).$$

A's rationality implies that she plays In if and only if she beliefs that B will play *Share* with probability greater that $\frac{1}{2}$. Therefore the set of states that survive the first iteration is $P_A^1 = C_A^1 \cup S_A^1$, where

$$C_A^1 = \left((In,\cdot),\alpha_A(In)=1,\alpha_A(Sh)\geq\frac{1}{2}\right)^{8},$$

$$S_A^1 = \left((Out),\alpha_A(In)=0,\alpha_A(Sh)<\frac{1}{2}\right).$$

The expected utility of B given that A played *In* is

$$\alpha_B(Sh|In)\Big(10-\theta_B^I(10m-10)\Big)+\alpha_B(Tk|In)\Big(14-\theta_B^S(14-0)\Big).$$

The rationality of B, associated with his preferences for equity, implies that B plays *Share* if his superiority sensitivity parameter is higher than the threshold $\hat{\theta}_m^S = \frac{2+5(m-1)\theta_B^I}{7}$. The set of states that survive the first iteration for player B is $P_B^1 = C_B^1 \cup S_B^1$, where

$$C_B^1 = \left((In,Sh),\theta_B^S\in[\hat{\theta}_m^S,\overline{\theta}_B^S],\alpha_B(Sh|In)=1\right),$$

$$S_B^1 = \left((In,Tk),\theta_B^S\in[\underline{\theta}_B^S,\hat{\theta}_m^S),\alpha_B(Sh|In)=0\right).$$

This means that for player A is rational to play *In* if and only if she believes that B will play *Share* with probability greater than 1/2. While player B chooses his action according to his type. Low superiority averse B plays *Take* and high superiority averse B plays *Share*.

[8]The ties are always broke in favor of the cooperative action.

2. Since A strongly believes P_B^1, $\alpha_A(Sh)$ must be equivalent to $\alpha_A(\theta_B^S \geq \hat{\theta}_m^S)$ since *Share* is consistent with B's rationality if and only if his sensitivity to superiority aversion is higher than the threshold $\hat{\theta}_m^S$. Therefore, $P_A^2 = C_A^2 \cup S_A^2$

$$C_A^2 = \Big((In, \cdot), \alpha_A(In) = 1, \alpha_A(\theta_B^S \geq \hat{\theta}_m^S) \geq \frac{1}{2}\Big),$$

$$S_A^2 = \Big((Out), \alpha_A(In) = 0, \alpha_A(\theta_B^S \geq \hat{\theta}_m^S) < \frac{1}{2}\Big).$$

Player B strongly believes in A's rationality and he anticipate that she plays *In* if and and only if she believes that B will play *Share* with probability higher that 1/2. Therefore the only conditional second order belief of B consistent with his strong belief in P_A^1 is $\beta_B(Sh|In) \geq \frac{1}{2}$. This step eliminates every state with second order beliefs inconsistent with A's rationality, therefore $P_B^2 = C_B^2 \cup S_B^2$, where

$$C_B^2 = \Big((In, Sh), \theta_B^S \in [\hat{\theta}_m^S, \overline{\theta}_B^S], \alpha_B(Sh|In) = 1, \beta_B(Sh|In) \geq \frac{1}{2}\Big),$$

$$S_B^2 = \Big((In, Tk), \theta_B^S \in [\underline{\theta}_B^S, \hat{\theta}_m^S), \alpha_B(Sh|In) = 0, \beta_B(Sh|In) \geq \frac{1}{2}\Big).$$

The result of this second iteration is that A plays *In* if and only if she beliefs that B has high superiority aversion with probability greater than 1/2. She chooses the cooperative action only if she is trustful enough.

3. The last iteration doesn't eliminate any state for player A, therefore $P_A^3 = P_A^2$. While player B, strongly believing P_A^2, recognizes that A recognized that, for B, *Share* is consistent only with $\theta_B^S \geq \hat{\theta}_m^S$. Therefore his conditional second order belief $\beta_B(Sh|In)$ must be equivalent to $\beta_B(\theta_B^S \geq \hat{\theta}_m^S|In)$. This implies $P_B^3 = C_B^3 \cup S_B^3$, with

$$C_B^3 = \Big((In, Sh), \theta_B^S \in [\hat{\theta}_m^S, \overline{\theta}_B^S], \alpha_B(Sh|In) = 1, \beta_B(\theta_B^S \geq \hat{\theta}_m^S|In) \geq \frac{1}{2}\Big),$$

$$S_B^3 = \Big((In, Tk), \theta_B^S \in [\underline{\theta}_B^S, \hat{\theta}_m^S), \alpha_B(Sh|In) = 0, \beta_B(\theta_B^S \geq \hat{\theta}_m^S|In) \geq \frac{1}{2}\Big).$$

Monotonicity Properties: Here I assume that there are populations of players A and B, and in each game the players are randomly selected from these populations. These populations are heterogeneous in personal traits and beliefs. I don't make any distributional assumption on these populations, beside assuming that these distributions are continuous and they have full support. Moreover I assume that all players have the maximum level of sophistication necessary, therefore they have beliefs consistent with P_i^3. These assumptions are a fair representation of the situation encountered during a lab experiment.

The fraction of players that play the cooperative action, respectively *In* and *Share*, is proportional to the size of the set C_i^3. First, lets focus on the threshold $\hat{\theta}_m^S$. This threshold is the only element that directly depends on the game parameter m. This threshold is equal to $\frac{2+5(m-1)\theta_B^I}{7}$ and it is clearly monotone increasing with m. Fix $m, m' \in \mathbb{R}$, with $m' > m \geq 1$. It is obvious that $[\hat{\theta}_{m'}^S, \overline{\theta}_B^S] \subseteq [\hat{\theta}_m^S, \overline{\theta}_B^S]$, which is equivalent to say $[\underline{\theta}_B^S, \hat{\theta}_m^S) \subseteq [\underline{\theta}_B^S, \hat{\theta}_{m'}^S)$. This implies that $S_B^3(m) \subseteq S_B^3(m')$, therefore all Bs that play *Take* in the game with parameter m they also play *Take* in the game with m'. Since C_B^3 and S_B^3 partition P_B^3 this means that the fraction of players B that play *Share*, $C_B^3(m)$, is weakly decreasing with m. The choice of player A is based on her beliefs about B's actions. Since B's actions are pinned down by his personal traits, this implies that As play *In* if and only if they believe that more than half of Bs' population has a sensitivity parameter θ_B^S greater that the threshold $\hat{\theta}_m^S$. Given the monotonicity of $\hat{\theta}_m^S$ this belief becomes more restrictive when m increases. Since all As with $\alpha_A(\theta_B^S \geq \hat{\theta}_m^S) < \frac{1}{2}$ also have $\alpha_A(\theta_B^S \geq \hat{\theta}_{m'}^S) < \frac{1}{2}$, this means that $S_A^3(m) \subseteq S_A^3(m')$, proving that the fraction of player A that play *In*, $C_A^3(m)$, is weakly decreasing with m.

The monotonicity of the set $[\hat{\theta}_m^S, \overline{\theta}_B^S]$ influences also the beliefs of the players. Since the event $\theta_B^S \in [\hat{\theta}_{m'}^S, \overline{\theta}_B^S]$ is included in the event $\theta_B^S \in [\hat{\theta}_{m'}^S, \overline{\theta}_B^S]$, the probability of the first event is lower than the probability of the second. This implies that for every player A, $\alpha_A(\theta_B^S \geq \hat{\theta}_{m'}^S) \leq \alpha_A(\theta_B^S \geq \hat{\theta}_m^S)$, and for every player B, $\beta_B(\theta_B^S \geq \hat{\theta}_{m'}^S|In) \leq \beta_B(\theta_B^S \geq \hat{\theta}_m^S|In)$. This proves that also the beliefs are monotone decreasing with m, and this conclude my analysis of the properties of the inequity aversion model.

3.4 Guilt Aversion

3.4.1 The model

Guilt is a form of emotional distress and psychologists Baumeister, Stillwell & Heatherton's (1994) argue that guilt arises upon the infliction of harm, loss, or distress on a relationship partner. The authors also state that if people feels guilty for hurting their partners by failing their expectations, they will alter their behavior in order to avoid this emotional distress. Inspired by this definition, Battigalli & Dufwenbrg (2007), B&D from now on, developed a game theoretical model of guilt aversion and they studied how this feeling shapes strategic interactions. B&D's model uses the framework of psychological game theory, framework firstly developed by Geanakoplos, Pearce and Stacchetti (1989) and extended by B&D (2009). In a psychological game the utility of each player not only depends on the actions of the players, but also on their beliefs, beliefs on what others believe, and so on. The term psychological game is slightly improper because it's not a game different from the traditional ones, but the game is analyzed using a new tool that allow to incorporate beliefs into utility functions. Since Baumeister et al. stated that guilt derives from failing partner's expectations, a game in which players are guilt averse must be analyzed as a psychological game.

In this model a player suffers a disutility proportional to the disappointment of the other player. The disappointment, D_i, is defined as the difference between the payoff that the player expects and the payoff that he receive, formally

$$D_i[y_i, \alpha_i] = \max\left\{\mathbb{E}_{\alpha_i}[y_i] - y_i(z), 0\right\}.$$

Therefore the utility function of a guilt averse player is

$$u_i(z) = y_i - \theta_i^G \left(D_{-i}[y_{-1}, \alpha_{-i}]\right),$$

where θ_i^G is the player's sensitivity to guilt. The first-order α_{-i} is not observable by player i, therefore, when player i evaluates his expected utility he must form a belief about $-i$'s beliefs. Therefore i's expected utility will depend on his second-order belief.

3.4.2 Beliefs and Type structure

As I previously did when analyzing the inequity aversion model, here I assume that guilt aversion is the only relevant preference. Again, I assume that player A is selfish. Therefore I use role dependent guilt aversion as my working hypothesis. I call G, short for *Guilt*, the event in which A is selfish and B is guilt averse. Formally I define the event G as

$$\begin{aligned} G &= \left(\Theta_A^G \times \Theta_A^S \times \Theta_A^I\right) \times \left(\Theta_B^G \times \Theta_B^S \times \Theta_B^I\right) \\ &= (\{0\} \times \{0\} \times \{0\}) \times \left([\underline{\theta}_B^G, \overline{\theta}_B^G] \times \{0\} \times \{0\}\right), \end{aligned} \tag{4}$$

where $\underline{\theta}_B^G \geq 0$ and $\overline{\theta}_B^G \geq \frac{4}{5}$. Again I assume $CSB^2(G)$, or in words, I assume that B is guilt averse, A strongly believes that B is guilt averse and B strongly believes that A strongly believes that B is guilt averse. Same for the selfishness of A.

Now I can proceed and show that in games with higher m there is a larger fraction of players that choose the cooperative actions.

3.4.3 Analysis

Here I follow the same procedure used previously. I proceed with the algorithm that eliminates the utility-relevant states not consistent with strong rationalizability. Again I divide those states in two sets, the sets of cooperative and selfish actions respectively, and at the end I show the monotonicity properties of those two sets. In the following I show that there are some $\theta_B^G \in \Theta_B^G$ for which both *Share* and *Take* can be rational. Therefore I need to introduce two additional sets.

I define the sets of states with θ_B^G such that it pins down a unique action, regardless of B's beliefs. I call $\overline{C}_B^n \subseteq C_B^n$ (resp. $\overline{S}_B^n \subseteq S_B^n$) the set of states that survived the n^{th} iteration, for which *Share* (resp. *Take*) is the only rationalizable action. I also define two similar sets for player A. Let be $\overline{C}_A^n \subseteq C_A^n$ (resp. $\overline{S}_A^n \subseteq S_A^n$) the set of states that survived the n^{th} iteration, for which A's beliefs over Θ_B^G pin down an unique rational action, *In* and *Out* respectively.

Strong Rationalizability The first three iterations of strong rationalizability give the key predictions:

1. The first iteration of the algorithm eliminates the utility-relevant states that violate rationality. Again, A's expected utility is

$$\alpha_A(In)\Big(\alpha_A(Sh)\cdot 10m\Big) + \alpha_A(Out)\Big(5m\Big).$$

A's rationality implies that she will play In if and only if she beliefs that B will play *Share* with probability greater that $\frac{1}{2}$. Therefore the set of states that survive the first iteration is $P_A^1 = C_A^1 \cup S_A^1$, where

$$C_A^1 = \left((In,\cdot), \alpha_A(In) = 1, \alpha_A(Sh) \geq \frac{1}{2}\right),$$

$$S_A^1 = \left((Out), \alpha_A(In) = 0, \alpha_A(Sh) < \frac{1}{2}\right).$$

Notice that the set P_A^1 is the equivalent to the set found in the previous analysis, the reason is that A's preferences haven't changed. In this model B suffers a disutility proportional to A's disappointment. If A plays *In*, she expects to earn $\alpha_A(Sh)\cdot 10m$. If B plays *Take*, she will earn 0, therefore her disappointment at the terminal history (In, Tk) will be equal to $\alpha_A(Sh)\cdot 10m$. Therefore the expected utility of B is equal to

$$\alpha_B(Sh|In)10 + \alpha_B(Tk|In)\Big(14 - \theta_B^G \cdot \beta_B(Sh|In)\cdot 10m\Big).$$

The best reply of B is not only determined by his sensitivity parameter, but also by his second-order belief. The action *Share* is B's best reply if and only if

$$\theta_B^G \geq \frac{4}{\beta_B(Sh|In) \cdot 10m}, \tag{5}$$

namely if his guilt $\theta_B^G \cdot \beta_B(Sh|In) \cdot 10m$ is greater than 4, which is the gain he gets if he chooses *Take* over *Share*.

Notice that if B has a sensitivity lower than $\frac{4}{10m}$, he will play *Take* regardless of his belief, since his belief can be at most equal to 1 and his guilt wouldn't be high enough. Therefore

$$\overline{S}_B^1 = \left((In, Tk), \theta_B^G \in [\underline{\theta}_B^G, \frac{4}{10m}), \alpha_B(Sh) = 0\right) \subseteq S_B^1.$$

This first step doesn't eliminate other states for player B because for every $\theta_B^G \geq \frac{4}{10m}$ and for every actions available to B, there exists some belief $\beta_B(Sh|In)$ the makes that action rationalizable.

2. Since A strongly believes in B's rationality, she realize that the least guilt averse players, namely those whit $\theta_B^G < \frac{4}{10m}$, will always play *Take*. Therefore for the less trustful As, those with $\alpha_A(\theta_B^G \geq \frac{4}{10m}) < \frac{1}{2}$, must be $\alpha_A(Sh) < \frac{1}{2}$, and *Out* is the only action consistent with rationality and her beliefs. Therefore

$$\overline{S}_A^2 = \left((Out), \alpha_A(In) = 0, \alpha_A(\theta_B^G \geq \frac{4}{10m}) < \frac{1}{2}\right).$$

This second iteration eliminates all the states with $z = (In, \cdot)$ for all players A that believe, with probability greater than $\frac{1}{2}$, that player B has a very low guilt sensitivity.

Since B strongly believes P_A^1, he knows that A plays *In* if and only if $\alpha_A(Sh) \geq \frac{1}{2}$, therefore it must be $\beta_B(Sh|In) \geq \frac{1}{2}$. This implies that all Bs with a sensitivity to guilt higher than $\frac{4}{5m}$ will always play *Share*. The threshold value $\frac{4}{5m}$ is obtained by substituting $\beta_B(Sh|In) = \frac{1}{2}$ in (5), which is the lowest value consistent with B's strong belief in P_A^1. This step doesn't eliminate other states for the low guilt types, those with $\theta_B^G < \frac{4}{10m}$, therefore I obtain

$$\overline{C}_B^2 = \left((In, Sh), \theta_B^G \in [\frac{4}{5m}, \overline{\theta}_B^G], \alpha_B(Sh) = 1, \beta_B(Sh|In) \geq \frac{1}{2}\right),$$

$$\overline{S}_B^2 = \left((In, Tk), \theta_B^G \in [\underline{\theta}_B^G, \frac{4}{10m}), \alpha_B(Sh) = 0, \beta_B(Sh|In) \geq \frac{1}{2}\right).$$

Notice that for all Bs with a sensitivity $\theta_B^G \in [\frac{4}{10m}, \frac{4}{5m})$, both *Share* and *Take* can be consistent with rationality and strong belief in rationality. If B's belief is greater than $\hat{\beta}_B(Sh|In) = \frac{4}{10m\theta_B^G}$, then his guilt will be high enough, therefore he will play *Share*. Otherwise he will play *Take*.

3. Since A believes P_B^2 she realize that the most trustworthy Bs will always play *Share* upon observing *In*. The most trustful As, those with $\alpha_A(\theta_B^G \geq \frac{4}{5m}) \geq \frac{1}{2}$, believe that the probability of finding a B with high guilt sensitivity, $\theta_B^G \in [\frac{4}{5h}, \overline{\theta}_B^G]$, is greater than $\frac{1}{2}$. Therefore those As believe $\alpha_A(Sh) \geq \frac{1}{2}$ and they will always play *In*. Therefore $\overline{C}_A^3$ is

$$\overline{C}_A^3 = \left((In, \cdot), \alpha_A(In) = 1, \alpha_A(\theta_B^G \geq \frac{4}{5m}) \geq \frac{1}{2}\right),$$

while $\overline{S}_A^3 = \overline{S}_A^2$

$$\overline{S}_A^3 = \left((Out), \alpha_A(In) = 0, \alpha_A(\theta_B^G \geq \frac{4}{10m}) < \frac{1}{2}\right).$$

Since B believes P_A^2, he realizes that A plays *In* if she trusts B enough, namely if $\alpha_A(\theta_B^G \geq \frac{4}{10m}) \geq \frac{1}{2}$. Beside this necessary condition on B's second-order beliefs, the sets $\overline{C}_B^3$ and $\overline{S}_B^3$ are the same of those of the previous iteration.

$$\overline{C}_B^2 = \left((In, Sh), \theta_B^G \in [\frac{4}{5m}, \overline{\theta}_B^G], \alpha_B(Sh) = 1, \beta_B(\theta_B^G \geq \frac{4}{10m}|In) \geq \frac{1}{2}\right),$$

$$\overline{S}_B^2 = \left((In, Tk), \theta_B^G \in [\underline{\theta}_B^G, \frac{4}{10m}), \alpha_B(Sh) = 0, \beta_B(\theta_B^G \geq \frac{4}{10m}|In) \geq \frac{1}{2}\right).$$

Monotonicity Properties: As I did when I analyzed the inequity aversion model I assume that there are two populations of players A and B. These population are heterogeneous in personal traits and beliefs about personal traits. Again I assume that the players are sophisticated enough to make 3 steps of iterated elimination of utility-relevant states.

The monotonicity analysis for the guilt aversion model is more challenging than the previous analysis. This is due to the fact that the second-order belief of B enter in his utility function, therefore his choices are not always uniquely pinned down by his guilt sensitivity. Therefore I'm not able to make clear cut predictions, but I'm still able to show that the cooperation's lower and upper bounds have monotonicity properties without making additional assumptions.

The fraction of players B that given their type only play *Share* is proportional to the measure of the set $\overline{C}_B^3$. Moreover it is also represent the minimum fraction of Bs that play *Share*. The measure of the set $\overline{C}_B^3$ is determined by the measure of the set $[\frac{4}{5m}, \overline{\theta}_B^G]$. Let fix $m, m' \in \mathbb{R}$, with $m' > m \geq 1$, then the following relationship holds $[\frac{4}{5m}, \overline{\theta}_B^G] \subseteq [\frac{4}{5m'}, \overline{\theta}_B^G]$. This relationship follows naturally given the monotonicity of the threshold $\frac{4}{5m}$ with respect to m. This shows that the minimum possible fraction of players B that play *Share* increases with m. Similarly the maximum fraction of *Share* is inversely proportional to the set $\overline{S}_B^G$, which is itself proportional to the measure of the set $[\underline{\theta}_B^G, \frac{4}{10m})$. Therefore the maximum fraction of *Share* also increases with m. Since both the minimum and maximum fraction of *Share*, it is reasonable to expect higher cooperation in a game with an higher m.

Since the choice of player A is based on her beliefs about B's type, the fraction of players A that play *In* inherit the monotonicity properties from the inclusion relationship of the sets $[\frac{4}{5m}, \overline{\theta}_B^G]$ and $[\underline{\theta}_B^G, \frac{4}{10m})$. The fraction of players A with $\alpha_A([\frac{4}{5m}, \overline{\theta}_B^G]) \geq \frac{1}{2}$ is the maximum possible fraction of players that play *In*. Since $[\frac{4}{5m}, \overline{\theta}_B^G] \subseteq [\frac{4}{5m'}, \overline{\theta}_B^G]$ for every $m, m' \in \mathbb{R}$, with $m' > m \geq 1$, $\alpha_A([\frac{4}{5m}, \overline{\theta}_B^G]) \geq \frac{1}{2}$ implies $\alpha_A([\frac{4}{5m'}, \overline{\theta}_B^G]) \geq \frac{1}{2}$. This shows that the maximum fraction of players that play *In* increases with m. Using a similar reasoning it is possible to show that also the minimum fraction of *In* is increasing in m.

Given inclusion relations of the sets $[\frac{4}{5m}, \overline{\theta}]$ and $[\underline{\theta}_B^G, \frac{4}{10m})$ is possible to show that minimum and maximum first and second-order beliefs increases with m. For a given m, $\alpha_A([\frac{4}{5m}, \overline{\theta}_B^G])$ is the highest first-order belief of A about B playing *Share*. The fact that $[\frac{4}{5m}, \overline{\theta}_B^G] \subseteq [\frac{4}{5m'}, \overline{\theta}_B^G]$ implies $\alpha_A([\frac{4}{5m'}, \overline{\theta}_B^G]) > \alpha_A([\frac{4}{5m}, \overline{\theta}_B^G])$. Same holds for A's lowest first-order belief $\alpha_A([\frac{4}{10m}, \overline{\theta}_B^G])$ and for B's highest and lowest second-order beliefs.

3.5 Comments

In this section I showed that the models give widely different predictions in some instances. These results not only give a theoretical foundation to my experimental design, but they also show that the two models are fundamentally different, although they are both used to explain the same evidences. The two models may offer the same predictions in those situations in which the payoffs of the two players are comparable in magnitude, but they tell two completely different stories when the stakes of one party are considerably higher. This implication should not be overlooked. In many real life situations happens that one party has significantly higher stakes in an investment project, and more often than not, it is the trustor. Therefore as economists we must have a clear understanding of the personal dynamics involved in these situations, since the implications of the two model are so deeply different.

One may dispute the implications of my findings by saying that I am comparing apples to oranges, because in the guilt aversion model there is only an altruistic component, while in the inequity aversion model there are both an altruistic and a resentful component.[9] I want to spend few moments addressing this possible criticism. Although this criticism is legitimate, it doesn't undermines completely my findings. In order to make the two models more comparable I can proceed in two ways. I can ignore the resentful component of the inequity aversion model, or I can add a resentful component to the guilt aversion model. Here I discuss both approaches. Using the first approach, the guilt aversion model offers the same predictions that I have analyzed previously, while the inequity aversion model offers different ones. Using only the altruistic component is easy to notice that now the utility of B is completely independent from m. Now the model predicts no differences in cooperation between the case with high and low m. The two models still offer different predictions, but now they are less pronounced. Although this approach is acceptable from a theoretical point of view, it is in contrast with many experimental evidences that show that inferiority aversion is higher than the superiority one. With the second approach I have to choose a resentful preference to add to guilt aversion. It makes sense to choose another belief dependent preference, and frustration is a good candidate. It turns out that B does not experience any frustration when A plays *In*, because the lowest possible material payoff for B is given by *Out*, therefore the frustration of B is zero. The last approach is to combine the two models, potentially substituting the superiority aversion with guilt. It is more difficult to get clear predictions from this mixed model. Both guilt and inferiority aversion grow linearly with m, but in opposite directions, possibly canceling each others out. A small or null effect of the treatment could actually suggest a mixed model, and, as many could expect, it is the case.

[9]With altruistic I mean that the material payoff of the coplayer enters positively in the utility function, while with resentful I mean that the coplayer's payoff enters in the utility function with a negative sign. I use the term resentful not only because the higher payoff of the coplayer hurts the player, but also because if this disutility is high enough it will resort in an action that punish the coplayer.

4 The Experiment

I this section I present my experiment. First, I describe the three treatments in detail. Next I outline the experimental procedures.

4.1 Experimental Design

The game played by the participants is a Trust Minigame. In this game one participant plays in the role of the trustor and another plays in the role of the trustee, I refer to these roles as A and B respectively. I present three different variations of the game and, from now on, the different treatments will be indicated using the parameter m employed. I perform a payoff manipulation between the three games. In the treatment with $m = 1.5$, the payoffs of A in every terminal history are increased by 50% with respect to the treatment with $m = 1$. While in the treatment with $m = 2$ the payoffs of A are doubled with respect the initial situation. Now I present the rules of the game for the treatment $m = 1$. The rules for the other two treatments are the same, with the only mentioned difference. Figure 3 shows the game forms of the three games.

[Figure 3 here]

A can choose to play *Out*, if she does, the game ends and both players earn 5€. Otherwise, she can choose to trust B and play *In*. Now, B can repay the trust that A put in him by playing *Share*, in this case both players earn 10€. Alternatively B can play *Take*, earning 14€, but leaving A with 0€. A and B make this choice at the same time. The Trust Minigame is a dynamic game with perfect information, in which A makes her choice and then B observes the choice made by A, and in the case in which A played *In*, B can then play *Share* or *Take*. Since in this experiment A and B make their choices simultaneously it is said that they are playing using the strategy method, meaning that they elicit the actions that are going to play when each contingency is met. I took care in explaining, both in the instructions and on the screen during the experiment, that the choice made by B is relevant only if A chooses to play *In*. The choice of using the strategy method is perfectly legitimate given the goal of this experiment, which is to investigate if the predictions given by the theory are accurate or not. This approach is theoretically justified because the preferences that I'm studying, guilt and inequality aversion, do not show dynamic inconsistency[10]. Moreover the strategy method has been used other experiments on trust,

[10] See Battigalli and Dufwenberg 2009 and Battigalli et al. 2019 for more details.

among those Charness and Dufwenberg (2006), Vanberg (2008) and the other papers previously cited.

Many will have already noticed that the material payoffs in the $m = 1$ treatment and the design are very similar to C&D (2006). The most notable difference between my design and C&D's is that I removed the chance move when B plays *Share*, *Roll* as is called in C&D. I made this decision because I want to study simple guilt and not other belief-based guilt models[11]. Thanks to this choice I am able to make clean predictions of the treatment effect, avoiding other confounding preferences, since the presence of the chance move, or hidden actions, triggers also other emotional responses in the subjects and the related preferences are needed in order to explain those emotions. Among those emotion there is shame, which is an emotion that can arise in a trust game, and Tadelis (2007) studied its effects in a Trust Minigame, both theoretically and experimentally.

With my design I am able to test two competing models, guilt and inequity aversion, one against the other. Now I explain the intuition behind my design. The more detailed and rigorous theoretical predictions are presented in the next section and they are derived from the analysis of the previous section.

All the experiments on Trust Games show that a positive and non negligible fraction of B respondents doesn't choose the selfish action *Take*. Similarly, a positive fraction of As chooses to trust B by playing *In*. These findings pose a challenge to standard game theory and to the assumption that players are selfish material payoff maximizers. These evidences also suggest that other regarding preferences, namely preferences in which the material payoff of the coplayer enters in the utility function, are needed to explain the behavior observed in the laboratory. Inequity and guilt aversion are two competing models that can explain these findings. The inequity aversion model predicts that B suffers a disutility when he plays *Take*. According to the inequity aversion model people dislike earning more or less than others. In the $m = 1$ game, if B plays *Take* he suffers a disutility caused by his superiority aversion equal to $\theta_B^S(14 - 0)$, where θ_B^S is his sensitivity to superiority aversion. If B's sensitivity is high enough he prefers to play *Share*. A is aware of B's preferences, and if she beliefs that his sensitivity is high enough she may choose to play *In*[12]. Therefore inequity aversion model successfully explains the experimental findings. In a similar fashion, also guilt aversion model can explain the same findings. If B is guilt averse he suffers a disutility when he play *Take*. This disutility is proportional to A's disappointment. A plays

[11] See Battigalli and Dufwenberg (2007) for other models, such as guilt from blame.

[12] As assumed in the previous Section ,A is selfish.

In if she expects B to play *Share* with probability greater than 1/2, and she expects to gain $\alpha_A^{Sh} \cdot 10$, where α_A^{Sh} is the first-order subjective belief of A that B plays *Share*. In the terminal history (*In*, *Take*), A's disappointment is equal to the difference between her expected and actual payoff, which is equal to $\alpha_A^{Sh} \cdot 10 - (1 - \alpha^{Sh}) \cdot 0$. Therefore the disutility suffered from B by playing *Take* is $\theta_B^G \cdot \beta_B^{Sh} \cdot 10$, where θ_B^G is B's sensitivity to guilt and β_B^{Sh} is his second order belief, namely his belief about A's belief about B playing *Share*. If θ_B^G and β_B^{Sh} are high enough, B prefers to play *Share* and A anticipates that and she plays *In*. Given the similar predictions of the two models, it is hard to understand which model drives the players' decisions in a Trust Minigame. The main difference between the two models is that guilt aversion is a belief dependent preference and there is positive correlation between B's beliefs and his actions. Higher B's second order beliefs lead to a higher frequency of *Share*. Since in the inequity aversion model there is no correlation between B's beliefs and actions, the presence of this correlation in experiments should be a proof in favor of the guilt aversion model. This is the reasoning used in previous papers.

With my design I am be able to give a stronger proof in favor of one of the two models, because they have opposite predictions for the treatment effect. The material payoffs of A are now increased by 50% and 100% in the treatments. This means that, in the inequity aversion model, B suffers a disutility from his inferiority aversion also in the terminal histories (*Out*) and (*In*,*Share*). Notice that B's inferiority aversion was not present in the $m = 1$ game and B's material payoffs are unchanged. Compared to the $m = 1$l game, now *Share* became less attractive for B because his utility now is $10 - \theta_B^I(15 - 10)$ for $m = 1.5$ and $10 - \theta_B^I(20 - 10)$ for $m = 2$, where θ_B^I is B's inferiority sensitivity. Now, in order for B to play *Share* he needs to have an higher θ_B^S to compensate for the disutility caused by his inferiority aversion, which means that only the subjects with the highest sensitivity, or more altruistic, will play *Share*. This implies that, according to the inequity aversion model, the frequency of *Share* in the treatment $m = 1.5$ and $m = 2$ will be lower than in $m = 1$l. The same holds for the frequency of *In*, since A anticipates B's reasoning.

The opposite is instead predicted by guilt aversion model. Since A's payoffs has been increased, now she expects to gain $\alpha^{Sh} \cdot 15$ and $\alpha^{Sh} \cdot 20$, respectively, by playing *In*. If B lets A down her disappointment will be greater, therefore also B's guilt will be increased proportionally. This implies that there will be an additional fraction of B subjects that would play *Share* in the last two treatments. Once again, this behavior is anticipated by A leading to a similar increase of *In*.

4.2 Experimental procedures

I recruited 120 participants from the subject pool of the university of Nice using ORSEE (Greyner 2015). The subjects pool includes students from many disciplines. The experiment was programmed using zTree (Fishbacher 2007) and run at the Laboratoire d'Economie Expérimentale de Nice (LEEN). The average pay was 14.70€, including the 5€ show up fee, and the experiments lasted on average 30 minutes. I ran 15 sessions[13] with 8 participants each, and each participant only played one of the three treatment. To have a comparable number of observation I evenly divided the sessions between the three treatments. Therefore I have 20 subjects that played in role A and 20 that played in role B for each treatment.

The first run of experiments was conducted between May 28^{th} and June 2^{nd} 2020 during the second phase of the Covid lockdown. During that time period there were in act stringent rules in order to maintain social distancing, therefore there was a limit of 8 people inside the laboratory. Thus I had to put a limit to the number of playable rounds in order to implement the desired matching protocol. To ensure everybody's safety masks were mandatory inside the lab, moreover every workstation was disinfected before the start of every experimental session. The original design prescribed only two treatments, with $m = 1$ and $m = 1.5$, but due to the absence of clear cut results, I decided to run additional sessions with $m = 2$. My hope was to see a change in the results, since in this new treatment the effect of the payoff manipulation is stronger. These additional sessions were conducted on April 20^{th} 2021. Although the Covid situation was slightly relaxed, these sessions were run using the same security protocols.

At the beginning of the experiment each participant was randomly assigned to either role A or role B, therefore in each session there was 4 players A and 4 players B, and the participants played in the same role for the entirety of the experiment. In each session the participants played 4 rounds, in each round participants in role A were matched with participants in role B according to an absolute typed stranger matching, meaning that each participant in a given role played once against every other participant in the other role. After the participants took place at their computer I read aloud the instructions and I made sure that the rules were understood. Moreover the instructions were printed and delivered to each subject.

In each round the participants faced two tasks and those tasks were the same during each of the four rounds. They had to choose the action to play

[13] There was another session in which I encountered a problem with the software and the data of the last round were not recorded. Due to this problem I will exclude the data of that session from the analysis.

and to elicit their belief and both task were displayed on the screen at the same time. In the decision task each player had to choose which action to play. This choice was different for the two roles, As had to choose between *In* and *Out*, while Bs had to choose between *Share* and *Take*.

In the belief elicitation I asked to the players in role A to elicit their first-order beliefs, namely to guess how many Bs will play *Share*. They had to choose between five options, between 0/4 and 4/4. To players B I asked them to elicit their second-order belies. They had to guess the guess made by their coplayer, thus they also had five options, from 0/4 to 4/4. The participants were paid according to the outcome of one of the rounds played, this round was randomly selected by the computer. Moreover they received 1€ each time that their belief was correct, except in the round that was selected for payment.

After the experiment, I asked to subject to fill a questionnaire. I asked them to report on a scale, from 0 to 10, their risk attitude, trust, guilt sensitivity, inferiority and superiority aversion sensitivity. Moreover I ask them to give their perception of the level of guilt, inferiority and superiority aversion of the general population. This questionnaire was not incentivized and all personal traits are self reported. The questionnaire was the same for all subjects, regardless their role during the game.

5 Behavioral Predictions

The main objective of this paper is to investigate the preferences of a trustee in order to better understand, or even predict, his behavior in real life situations that resemble a trust game. In Section 3 I analyzed the game twice, each time assuming that only one preference is relevant and there is consensus[14] about that. This assumption makes the analysis of each model easier and it delivers clear-cut results, but it is fairly unrealistic to expect that in the real world. Everyone in their personal life experiences both guilt and a distaste for inequality, therefore it is obvious to think that both models will affect the behavior of the participants of an experiment. Nonetheless, I try to test the predictions derived from the results of Section 3, therefore I assume and test the two most extreme cases. The predictions are derived by setting $m = 1$ for the control and $m = 1.5$ and $m = 2$ for the treatments. For every behavioral prediction I present two hypothesis, one that is implied by inequity aversion and the other that is implied by guilt aversion. These pairs of hypothesis are always mutually exclusive. If evidences in favor of a specific models is found, it won't be a proof that the other preference doesn't exist, but that one preference is more relevant in the context analyzed.

5.1 Behavioral predictions about the trustee

The first behavioral prediction that I test is the frequency with which players Bs choose *Share*. The prediction given by the inequity aversion model is obtained by simply assuming preferences for equity for player B and his rationality.

Hypothesis 1.a. If players B are inequity averse, then the frequency of *Share* is decreasing as m increases between treatments.

In order to give predictions of B's behavior, the guilt aversion model also requires that B strongly believes A's rationality. The guilt aversion model requires two steps of deletion of non-best replies, therefore it requires a higher strategical sophistication.

Hypothesis 1.b. If players B are guilt averse, then the frequency of *Share* is increasing as m increases between treatments..

[14] I use "consensus about a preference", for instance role-dependent guilt aversion, in an informal and more compact way, meaning that everybody strongly believes role-dependent guilt aversion and everybody strongly believes that everybody strongly believes role-dependent guilt aversion.

5.2 Behavioral predictions about the trustor

Next I compare the behavior of players A across treatments. The behavior of the trustor is not influenced by her preferences, but by her beliefs about B's preferences and rationality. In order to give predictions, both models require that player A strongly believes that player B is either inequity or guilt averse. Once again, the guilt aversion model requires a higher sophistication from player A. The predictions on the frequency of *In* across treatments given by, respectively, inequity and guilt aversion model are the followings.

Hypothesis 2.a. If players A strongly believe that players B are inequity averse, then the frequency of *In* is decreasing as m increases between treatments.

Hypothesis 2.b. If players A strongly believe that players B are guilt averse, then the frequency of *In* is increasing as m increases between treatments.

5.3 Predictions about elicited beliefs

During the experiment the subjects are asked to report their beliefs. Players A are asked to guess how many Bs will play *Share*, hence their first-order belief. While players B are asked to guess the guess made by A, their second-order belief. If consensus about a particular preference is assumed, the two models give predictions about the reported beliefs of the players. In order to make this predictions, the highest level of sophistication considered is required, hence three steps of deletion of non best reply.

Hypothesis 3.a. A's guess of the number of Bs playing *Share* is decreasing as m increases between treatments.

Hypothesis 3.b. A's guess of the number of Bs playing *Share* is increasing as m increases between treatments.

Hypothesis 4.a. B's guess of the guess made by A is decreasing as m increases between treatments.

Hypothesis 4.b. B's guess of the guess made by A is increasing as m increases between treatments.

6 Results

In this section I present the experimental results that I obtained. I start by reporting the results for B's behavior, then I report A's behavior and the beliefs of both roles. There were 120 players in total, equally divided among roles and treatments. This means that there were 20 subjects that played in role B and 20 that played in role A in each treatment. Each subject played the same game four times, therefore I have 80 data points for every role in every treatment.

Notice that in the first run of experiment there was only the treatments $m = 1$ and $m = 1.5$. Since the preliminary results were not promising, I decided to run additional experimental session with $m = 2$, hoping that with a stronger treatment variable I would find clearer results. For this reason I first present the comparison between treatments $m = 1$ and $m = 1.5$, where the first treatment acts as control. Then I will compare treatments $m = 1$ and $m = 2$, always treating $m = 1$ as control, and I will discuss eventual changes in the results. First I make a preliminary analysis of the data. I test the treatment effects using a Mann-Withney test. As stated before, the treatment with $m = 1$ is used as control, and the other two treatments are each used as treated group for their respective analysis. This test considers the 4 observations of each player as independent observations. I take care of this issue later, where I use panel data. Moreover in this more accurate analysis I will all three treatments together, using m as a continuous variable.

6.1 Preliminary Analysis

Here I discuss the preliminary results, first presenting the results obtained in the first run of experiment, then I analyze the new data from the second run and I compare the results. Figures 4 and 5 show the frequencies of the actions played in all three treatments. While the distributions of the elicited first and second-order beliefs are shown in Figure 6 and Figure 7, respectively.

[Figure 4 Here]
[Figure 5 Here]
[Figure 6 Here]
[Figure 7 Here]

Control Group: $m = 1$, **Treated Group:** $m = 1.5$ I start by testing the first prediction about the behavior of players B. In the control group 40% (32/80) of B subjects played *Share*, and 42.5% (34/80) played *Share* in the treated group. The frequency of *Share* is slightly higher in the treated group, but the difference is not statistically significant (Mann-Withney p-value 0.75). I can't accept hypothesis 1.a. (0.63 one-sided t-test), nor hypothesis 1.b. (0.38 one-sided t-test). The results do not support the assumption that one model is more relevant than the other in a trust game.

Now I investigate the behavior of A subjects. The frequency of *In* in the control group is 52.5% (42/80), while the frequency of *In* in the treated group is 32.5% (26/80). The difference between the two groups is statistically significant (MW p-value 0.01). The hypothesis 2.a. is accepted (one-sided t-test 0.005), while the hypothesis 2.b. is rejected (one-sided t-test 0.99). This result show that the subjects believe that Bs are inequity averse, although this fact is not confirmed by the experimental evidences.

The higher distrust of players A in the treated group is reflected also by their first-order beliefs. The average elicited first-order belief (on a 0-4 scale) in the control is 2.15, while in the treated group is 1.71. The difference is statistically significant (MW p-value 0.022) and hypothesis 3.a. in favor of inequity aversion is accepted, while hypothesis 3.b. must be rejected. The average second-order belief elicited by Bs in the control is 2.31, while it is 2.26 in the treated group. The difference is not statistically significant (MW p-value 0.89) and neither hypothesis 4.a., nor 4.b., can be accepted. Again this results suggest that As become less trustful when their stakes are higher, although this lack of trust is unjustified given the behavior of Bs.

Control Group: $m = 1$**, Treated Group:** $m = 2$ I proceed with replicating the same analysis done before, but using the treatment $m = 2$ as treated group. The frequencies of the actions played in the $m = 2$ treatment don't differ much from the previous $m = 1.5$ treatment. B subjects played *Share* 42.5% (34/80) of the times, exactly like in the treatment $m = 1.5$, despite the fact that the treatment effect should be stronger. Therefore, as before, there is no significant difference between the treated and control group (Mann-Withney p-value 0.75).

Similar is the situation for the A subjects that played *In* 31.25% (25/80), one less than in treatment $m = 1.5$. As before, The difference between the two groups is statistically significant (MW p-value 0.007). The hypothesis 2.a. is accepted (one-sided t-test 0.003), while the hypothesis 2.b. is rejected (one-sided t-test 0.99).

Regarding the first order beliefs, the distrust of A players is still present, and even stronger, with an average first-order belief of 1.48, reconfirming the statistically significant difference (MW p-value 0.001) between the control an treated group. Once again, the results regarding the second-order beliefs are in line with those found in the first run of experiment, with an average second-order belief of 2.18 and no statistical difference with the control group (MW p-value 0.76).

Finally, for every relevant variable (choice and beliefs) there is no statistical difference between the treatments $m = 1.5$ and $m = 2$.

Final Remarks These results show an interesting picture. There is no evidence that the subjects are driven mainly by only one motivation. A possible explanation is that subjects face a tension between the two motivations, in the treatment the higher guilt experienced by the subject is balanced by their inferiority aversion. This suggests that a mixed model, one with both guilt and inequity aversion, could be better suited for predicting trustees behavior. Despite that, trustors fail to recognize the balance between the two motivations and they give more relevance to inferiority aversion. This lead to a significant lower trust in both treatments with an higher m. It is also puzzling the fact that there is no change, in either direction, when the treatment variable is increased from 1.5 to 2. The small number of participants is a primary concern when discussing the results, but nevertheless the small effect of the treatment variable is not promising.

6.1.1 Gender Effects

This study was not designed to test gender differences in the behavior of the subjects. I analyzed the behavior of males and females separately I found the different genders react differently to the treatment variable. In the control, $m = 1$, 10 male subjects participated in role of B. Those played *Share* 35% of the times (14/40). While there were 5 males in the role of B in the treatment $m = 1.5$ and they played *Share* 65% of the times (13/20). If only males are concerned the difference between the two initial treatments is significant (MW p-value 0.03) and relevant, since the frequency of *Share* almost doubled in the $m = 1.5$ treatment.

For the females it is not possible to draw the same conclusion. There is a decrease of the frequency of *Share* from the control to the treatment, from 45% (18/40) to 35% (21/60), but the difference is not significant (MW p-value 0.32). Moreover, the behavior of males and females in the control is comparable, with 35% of *Share* for males and 45% for females, the difference is not statistically significant. The same can't be said for their behavior in the treatment $m = 1.5$. Males cooperate significantly more that females, 65% and 35% respectively. The difference is statistically significant (MW p-value 0.02). This result suggest that females are not motivated by guilt as much as males and it seems that females are more concerned by equity. Despite these apparent differences there is no significant difference between the behavior of males and females in the control (MW p-value 0.36), their behavior differs significantly only in the treatment $m = 1.5$ (MW p-value 0.02). It is important to underline that in the treatment there are significantly more females that played in the role of B, therefore the sample was unbalanced with respect to this variables. This unbalance undermine the results, or lack of thereof, previously reported on the behavior of players in the role of B.

Indeed, if we consider now the $m = 2$ treatment, where there was 8 male subjects, we don't observe the same pattern. Males played *Share* 40.6% of the times (13/32), while females played *Share* 43.75% of the times (21/48), and there is no statistically significant difference between the two genders (MW p-value 0.78).

Instead, there is no significant difference between the behavior of males and females that played in the A role, and this fact is consistent across all three treatments.

6.2 Panel Regression

Here I make a more rigorous analysis accounting for the fact that each participant makes four choices during the experiment, and those choices should not regarded as independent. Moreover I use also the personal traits a demographics reported in the final questionnaire as independent variables. In the final questionnaire, along with demographics, I asked the participants to report some personal traits, and their opinion about those personal traits in the general population. The questionnaire can be found in the appendix. I ask them to report their risk attitude, trust, guilt, inferiority and superiority sensitivities, moreover I ask them to give their opinion about the guilt, inferiority and superiority sensitivities of others. I run panel regressions with random effects for each player role, those regressions are summarized in Table 1 and Table 2. The choice of using random effects is supported by the results of the Breusch-Pagan and Hausman tests.

For players A the variables that predict their choices are their beliefs and risk attitude. Higher the first-order beliefs of A, more likely she plays *In*, this in line with rationality. Risk attitude also plays a major role in the choice. More risk loving people plays *In* more often. This result requires a comment. In the treatment the choice *In* is more risky. I found that the frequency of *In* is lower in the treatment and I explained this finding stating that As believe in the inequity aversion of Bs. This intuition is not fully supported by the regressions. A lower frequency of *In* could be explained by the effect of the risk attitude on choices, moreover seems that the superiority aversion sensitivity perceived by As doesn't play any role in their choice. If risk attitude is the main driver of the behavior of A it could be interesting, and maybe necessary, to introduce a treatment that keeps constant the variance. It can be easily done by reducing the payoff of B instead of increasing the payoff of A. The result for players B are quite puzzling. None of the variables is strongly affecting the choices. It also puzzling the positive and significant effect of trust. B is the one who is being trusted, not the opposite. Moreover the self reported sensitivities play no role in his choice.

7 Conclusions

This paper presents a theory driven experiment to test the model of guilt aversion against the model of inequity aversion in the context of a Trust Minigame. The first contribution of this paper is theoretical. I showed that the two models under investigation react differently to payoff manipulations. The two models were treated independently without assuming complete information or equilibrium play. Moreover I highlighted that the predictions given by each models highly depended on the beliefs that players hold about the type structure and beliefs about those beliefs, while in other works those beliefs are taken for granted or even assume that players hold correct beliefs.

Indeed the experimental results show that players do not hold correct beliefs and they fail to foresee the motivations of their partner. This work fails to provide definitive evidences in favor of a model over the other, suggesting that guilt and inequity aversion can't be used a substitutes, but they integrate each other. I run additional experimental sessions with a stronger treatment variable, hoping to detect an effect in either direction. Also the new data confirm the same results, or lack of thereof, found in the first run.

It remains unclear if the failure of this experiment is due exclusively to the small sample, or instead is due to the limitation of the theories under investigation. It is undeniable that the number of participants is in the low end, but the results remain puzzling. The most notable is the fact that the results of the $m = 1.5$ and $m = 2$ treatments are fundamentally the same, when both theories predicts an effect. The evidences, or lack of thereof, pose some question regarding some assumptions of the two models, mostly the linearity of the payoffs. Further investigation, either experimental or theoretical, is needed.

References

[1] Pierpaolo Battigalli, Roberto Corrao, and Martin Dufwenberg. Incorporating belief-dependent motivation in games. *Journal of Economic Behavior & Organization*, 167:185–218, 2019.

[2] Pierpaolo Battigalli and Martin Dufwenberg. Guilt in games. *American Economic Review*, 97(2):170–176, May 2007.

[3] Pierpaolo Battigalli and Martin Dufwenberg. Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35, 2009.

[4] Pierpaolo Battigalli and Marciano Siniscalchi. Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106(2):356–391, 2002.

[5] Roy Baumeister, Arlene Stillwell, and Todd Heatherton. Guilt: An interpersonal approach. *Psychological bulletin*, 115:243–67, 04 1994.

[6] Joyce Berg, John Dickhaut, and Kevin McCabe. Trust, reciprocity, and social history. *Games and economic behavior*, 10(1):122–142, 1995.

[7] S Braver. Social contracts and the provision of public goods. *Social dilemmas: Perspectives on individuals and groups*, pages 69–86, 1995.

[8] Gary Charness and Martin Dufwenberg. Promises and partnership. *Econometrica*, 74(6):1579–1601, 2006.

[9] Gary Charness and Martin Dufwenberg. Bare promises: An experiment. *Economics letters*, 107(2):281–283, 2010.

[10] Ying Chen, Navin Kartik, and Joel Sobel. Selecting cheap-talk equilibria. *Econometrica*, 76(1):117–136, 2008.

[11] Giovanni Di Bartolomeo, Martin Dufwenberg, Stefano Papa, and Francesco Passarelli. Promises, expectations & causation. *Games and Economic Behavior*, 113:137–146, 2019.

[12] Giovanni Di Bartolomeo, Martin Dufwenberg, and Stefano Papac. Promises and partner-switch. 2020.

[13] Tore Ellingsen and Magnus Johannesson. Promises, threats and fairness. *The Economic Journal*, 114(495):397–420, 2004.

[14] Ernst Fehr and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868, 1999.

[15] Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178, 2007.

[16] John Geanakoplos, David Pearce, and Ennio Stacchetti. Psychological games and sequential rationality. *Games and economic Behavior*, 1(1):60–79, 1989.

[17] Ben Greiner. Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125, 2015.

[18] Navin Kartik. Strategic communication with lying costs. *The Review of Economic Studies*, 76(4):1359–1395, 2009.

[19] David M Kreps. Corporate culture and economic theory. *Perspectives on positive political economy*, 90(109-110):8, 1990.

[20] Elinor Ostrom, James Walker, and Roy Gardner. Covenants with and without a sword: Self-governance is possible. *American political science Review*, 86(2):404–417, 1992.

[21] Steven Tadelis. The power of shame and the rationality of trust. *Available at SSRN 1006169*, 2007.

[22] Christoph Vanberg. Why do people keep their promises? an experimental test of two explanations 1. *Econometrica*, 76(6):1467–1480, 2008.

A Instructions

A.1 Control

Thank you for participating in this session. The purpose of this experiment is to study how people make decisions in a particular situation. Feel free to ask us questions as they arise, by raising your hand. Please do not speak to other participants during the experiment.

You will receive €5 for participating in this session. You may also receive additional money, depending on the decisions made (as described below). Upon completion of the session, this additional amount will be paid to you individually and privately.

The experiment consists of 4 independent rounds. In each round, you will interact in a game with a different participant randomly chosen by the computer. You will never interact with the same participant more than once. No participant will even know the identity of the subject with whom he or she interacted during the experiment. Your payment will be based on the decisions you will make during one of the eight rounds. That round will be randomly selected at the end of the experiment (each round has the same probability of being selected).

At the beginning of the experiment you will randomly be assigned to either role A or role B (based on your role, you will be asked to perform different tasks). If you are assigned role A, your partner will take role B and vice versa. You will play the whole session in your assigned role.

During this decision phase, Participant A will choose *In* or *Out*. If A chooses *Out*, A and B each receive €5. Simultaneously B will choose *Share* or *Take* and B's decision will only make a difference when A chooses *In*. If A chooses *In* and B chooses *Share* they both receive €10. If A chooses *In* and B chooses *Take* B receives €14 while A receives €0. This Information is summarized in the chart below:

	Player A earns	**Player A earns**
If A chooses *Out*	€5	€5
If A chooses *In* and B chooses *Share*	€10	€10
If A chooses *In* and B chooses *Take*	€0	€14

Additional earnings: Moreover, at every round of the experiment, you will have the opportunity to earn additional money if you answer some questions correctly. A will be asked to guess how many B will play SHARE.

While B will be asked to guess the guess made by A with whom he is paired with. You will be paid €1 for each correct guess.

You will always be paid when your guesses are correct, except in the round that will be drawn by the computer for payment. In that round, you will not be paid for your guesses, but only for the outcome of the decision task.

A.2 Treatment

Thank you for participating in this session. The purpose of this experiment is to study how people make decisions in a particular situation. Feel free to ask us questions as they arise, by raising your hand. Please do not speak to other participants during the experiment.

You will receive €5 for participating in this session. You may also receive additional money, depending on the decisions made (as described below). Upon completion of the session, this additional amount will be paid to you individually and privately.

The experiment consists of 4 independent rounds. In each round, you will interact in a game with a different participant randomly chosen by the computer. You will never interact with the same participant more than once. No participant will even know the identity of the subject with whom he or she interacted during the experiment. Your payment will be based on the decisions you will make during one of the eight rounds. That round will be randomly selected at the end of the experiment (each round has the same probability of being selected).

At the beginning of the experiment you will randomly be assigned to either role A or role B (based on your role, you will be asked to perform different tasks). If you are assigned role A, your partner will take role B and vice versa. You will play the whole session in your assigned role.

During this decision phase, Participant A will choose *In* or *Out*. If A chooses *Out*, A receives €7.5 and B receive €5. Simultaneously B will choose *Share* or *Take* and B's decision will only make a difference when A chooses *In*. If A chooses *In* and B chooses *Share* A receives €15, while B receive €10. If A chooses *In* and B chooses *Take*, B receives €14 while A receives €0. This Information is summarized in the chart below:

	Player A earns	**Player A earns**
If A chooses *Out*	€7.5	€5
If A chooses *In* and B chooses *Share*	€15	€10
If A chooses *In* and B chooses *Take*	€0	€14

Additional earnings: Moreover, at every round of the experiment, you will have the opportunity to earn additional money if you answer some questions correctly. A will be asked to guess how many B will play SHARE. While B will be asked to guess the guess made by A with whom he is paired with. You will be paid €1 for each correct guess.

You will always be paid when your guesses are correct, except in the round that will be drawn by the computer for payment. In that round, you will not be paid for your guesses, but only for the outcome of the decision task.

A.3 Questionnaire

Socio-Demographics

- How old are you?
- What is your gender? Male Female
- What is your occupation?
  ☐ Student
  ☐ Employee
  ☐ Unemployed
  ☐ Retired
  ☐ Other
- If you are a student, what is your field of study?
  ☐ Economy and management
  ☐ Social Sciences
  ☐ Arts and Humanities
  ☐ Engineering Sciences
  ☐ Medical studies
  ☐ Other
- What is your level of study?
  ☐ Elementary school license
  ☐ Middle school license
  ☐ High school license
  ☐ Bachelor's degree
  ☐ Post-graduate degree
- How much experience have you had with LEEN before?

Psychological questions: First Part

- How do you see yourself : are you a persons that usually is prone to take risk?

  Choose your answer between 0 and 10, 0 means that you always avoid risk and 10 means that you always willing to take risk.

  0 1 2 3 4 5 6 7 8 9 10

- How do you see yourself : are you a persons that usually trust others?

  Choose your answer between 0 and 10, 0 means that you never trust others and 10 means that you always trust others.

  0 1 2 3 4 5 6 7 8 9 10

- How do you see yourself : are you a person that easily feels guilty?

  Choose your answer between 1 and 10, where 1 means that you don't easily feel guilty and 10 means that you very easily feel guilty.

  0 1 2 3 4 5 6 7 8 9 10

- How do you see yourself : are you a person that usually dislike to earn less than the others?

  Choose your answer between 1 and 10, where 1 means that you don't dislike it at all and 10 means that you absolutely hate earning less than others.

  0 1 2 3 4 5 6 7 8 9 10

- How do you see yourself : are you a person that usually dislike to earn more than the others?

  Choose your answer between 1 and 10, where 1 means that you don't dislike it at all and 10 means that you absolutely hate earning more than others.

  0 1 2 3 4 5 6 7 8 9 10

Psychological questions: Second Part

- What do you think about other people : do you think that usually people feel easily guilty or they feel hardly guilty?

  Choose your answer between 0 and 10, where 0 means that you think that other people hardly feel guilty and 10 means that you think that they feel easily guilty.

  0 1 2 3 4 5 6 7 8 9 10

- What do you think about other people :do you think that usually people dislike to earn less than the others?

  Choose your answer between 0 and 10, where 0 means that you think that other people don't dislike it at all and 10 means that you think that people absolutely hate earning less than the others.

  0 1 2 3 4 5 6 7 8 9 10

- What do you think about other people : do you think that usually people dislike to earn more than the others?

  Choose your answer between 0 and 10, where 0 means that you think that other people don't dislike it at all and 10 means that you think that people absolutely hate earning more than the others.

  0 1 2 3 4 5 6 7 8 9 10

B Tables and Figures

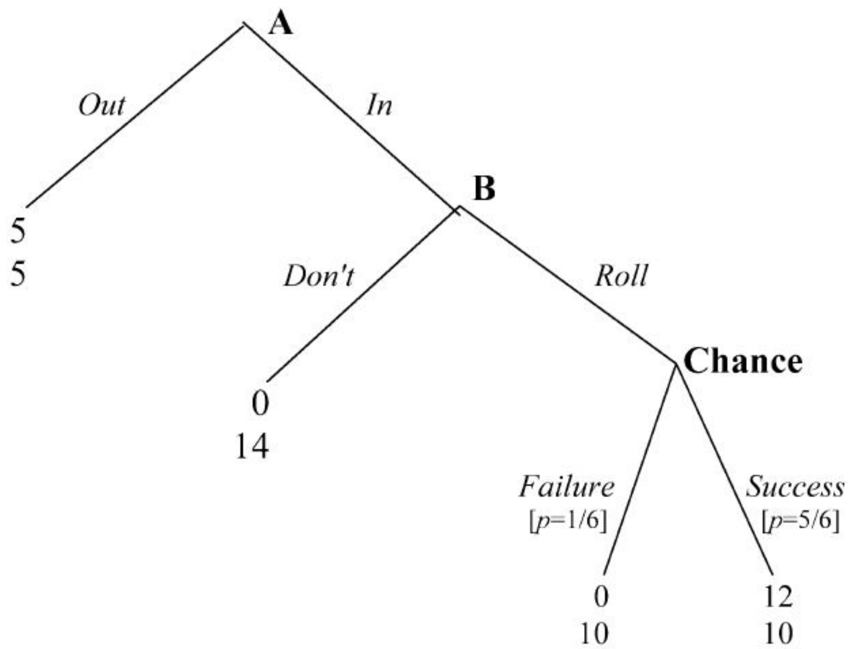

Figure 1: Game form used in Charness & Dufwenberg (2006).

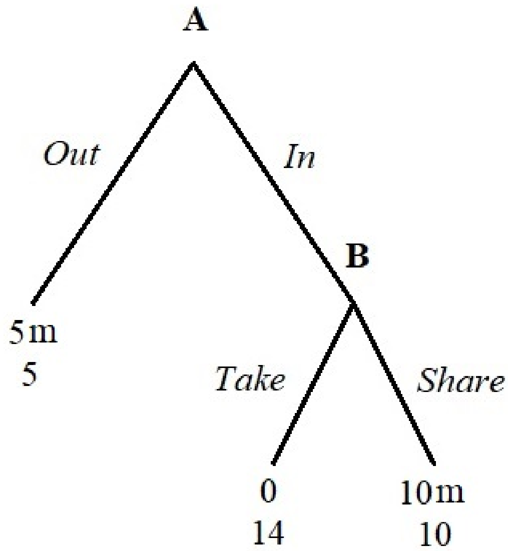

Figure 2: Parametrized version of the Trust mini-Game.

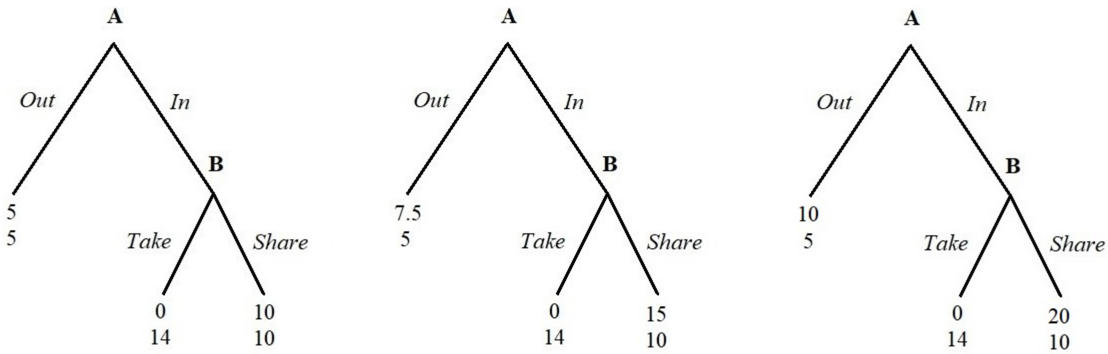

Figure 3: Trust Game game forms of the 3 treatments, with $m = 1$, $m = 1.5$ and $m = 2$ respectively.

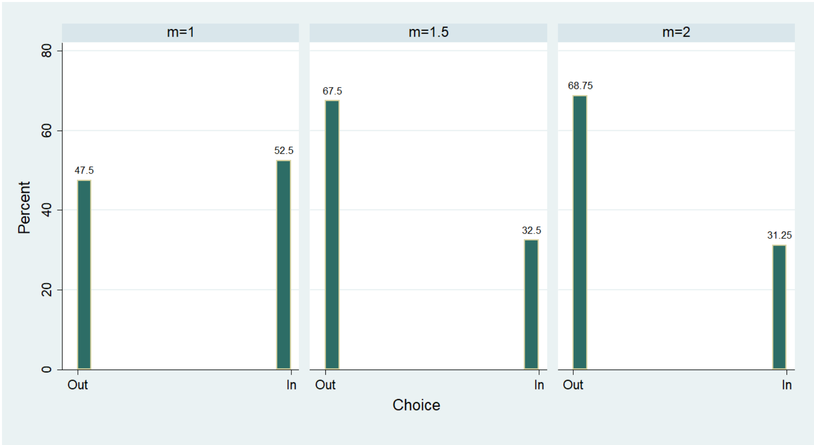

Figure 4: frequency of *In* in the three treatments.

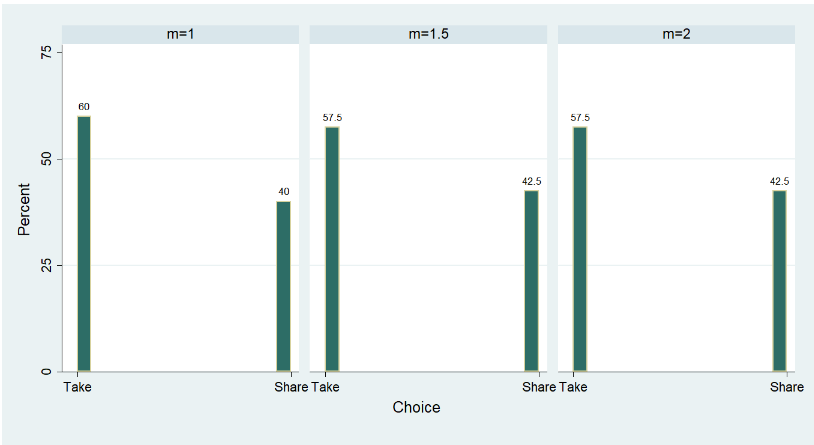

Figure 5: frequency of *Share* in the three treatments.

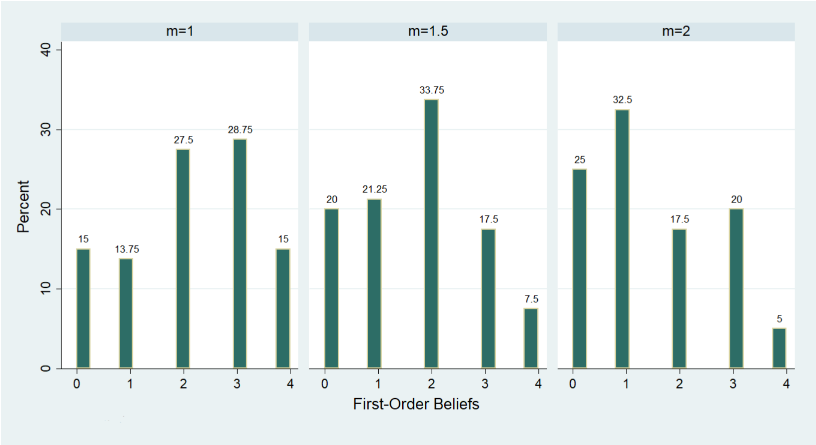

Figure 6: A's elicited first-order beliefs

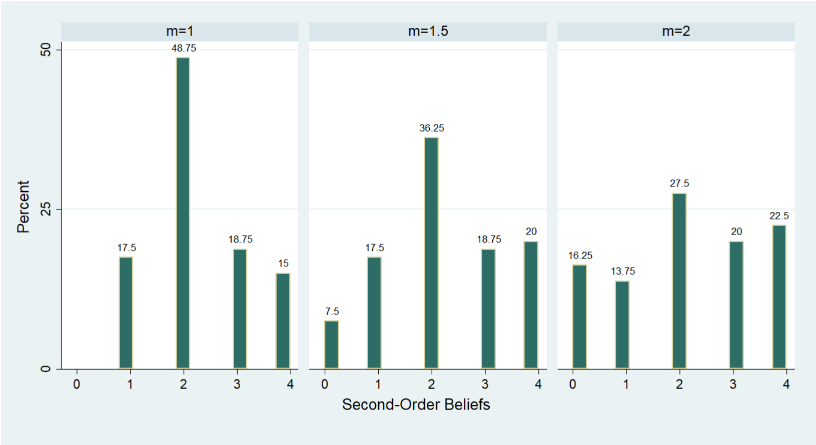

Figure 7: B's elicited second-order beliefs

	(1) Choice	(2) Choice	(3) Choice
Treatment (m)	-0.126	-0.125	-0.0915
	(-1.43)	(-1.47)	(-0.99)
Period	-0.0278	-0.0278	-0.0271
	(-1.14)	(-1.13)	(-1.11)
Belief	0.128***	0.128***	0.134***
	(5.04)	(5.20)	(5.39)
Sex		0.0537	0.0534
		(0.77)	(0.76)
Age		-0.000120	0.00115
		(-0.01)	(0.11)
Experience		-0.0145	-0.0147
		(-1.25)	(-1.23)
Risk		0.0603***	0.0596***
		(3.71)	(3.65)
Trust		0.000865	0.000308
		(0.05)	(0.02)
Guilt of Others			0.00515
			(0.26)
Inferiority of Others			-0.00262
			(-0.17)
Superiority of Others			-0.0241
			(-1.71)
N	240	240	240

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1: Panel Regressions: Role A

	(1) Choice	(2) Choice	(3) Choice
Treatment (m)	0.0228	0.0415	-0.00575
	(0.19)	(0.37)	(-0.05)
Period	-0.0454*	-0.0449*	-0.0450*
	(-2.19)	(-2.17)	(-2.18)
Belief	-0.0173	-0.0233	-0.0219
	(-0.72)	(-0.96)	(-0.90)
Sex		-0.0304	-0.0315
		(-0.29)	(-0.30)
Age		-0.0286*	-0.0304*
		(-2.36)	(-2.49)
Experience		0.0126	0.0114
		(0.61)	(0.55)
Risk		0.00423	0.00600
		(0.19)	(0.26)
Trust		0.0844***	0.0717**
		(3.92)	(3.01)
Guilt			0.0282
			(1.47)
Inferiority			-0.00773
			(-0.45)
Superiority			0.0164
			(0.88)
N	240	240	240

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Panel Regressions: Role B

Chapter 2

The Enemy of my Enemy

The Enemy of my Enemy: Competitive Framing in Repeated Prisoner's Dilemmas

Sara Gil-Gallen*[1] and Alessandro Stringhi†[2]

[1] *Università degli studi di Bari "Aldo Moro", Italy*
[2] *Università Commerciale Luigi Bocconi, Italy*

Abstract

The tension between selfish behavior and cooperation is a social dilemma often encountered in ordinary life which essence is captured by the Prisoner's Dilemma. This tension can be alleviated or exacerbated if the game is framed in a cooperative or competitive way, respectively. In this study, we investigate how a competitive framing can increase cooperation as long as the hostility is redirected from the partner to an opposing pair of players. We design an experiment in which we frame a infinitely repeated Prisoner's Dilemma as a competition between pairs of players whose goal is to accumulate more aggregate payoffs. The findings show that a simple framing device, even without additional monetary rewards, is able to increase cooperation among participants in a controlled lab experiment. Moreover, the effect gets stronger as the players gain experience, showing the power and reliability of this framing device.

1 Introduction

The social dilemma between selfish behaviors and cooperation emerges during day to day human interactions. This tension does not only affect human relationships, but it also affects individual behavior in a society, since there is a trade-off between what is socially, but not individually optimal. This social dilemma is of great interest and it has been investigated in so

*saragilgallen@gmail.com
†alessandro.stringhi@phd.unibocconi.it

many studies that an extensive review of the literature will go beyond the aim of this work[1].

In this paper, instead, we focus our attention to a particular behavior firstly reported in the seminal experiment of Deutsch (1958); how the framing of the instructions induces the subjects to behave in a cooperative or individualistic way in a Prisoner's Dilemma. In his work, the instructions were heavily loaded, strongly implying the behavior that was expected from the participants. Subsequent works show that simply changing the name of the game has significant effect on behavior. Eiser & Bhavnani (1974) finds more cooperation when the situation in the game is described as an international negotiation rather than a business transactions. Others studies (Kay & Ross, 2003; Liberman et al., 2004; Ellingsen et al., 2012) find the same differences between games called *"Community Game"* and *"Stock Market Game"* respectively, without relying on additional framing or context. While, the results found in Engel & Rand (2014) offers a different interpretation of the effect previously reported. In the paper, the authors find that there is no significant difference between games with a neutral or competitive framing, but there is a significant lower cooperation when the game is labeled in a competitive way.

The aim of the present paper is to expand the literature on framing effects in a Prisoner's Dilemma by designing and implementing a new experiment. Our goal is to understand if a competitive framing can be used to increase cooperation. To do so, it is necessary to deflect the rivalry from the partner to someone else, therefore we implement a tournament in which pairs of players compete with the goal of scoring more points, expressed as the sum of the payoffs of both members of the pair. Indeed, the competitive framing that we propose, as we call it, tournament, is different from previous works because the competition is deflected from the partner to an opposing pair of players. Moreover, in our design, the subjects will play a series of indefinitely repeated Prisoner's Dilemma, while in previous works the subject play a game, one-shot or finitely repeated, only once. By letting the subjects play more games allow us to study the evolution and sustainability over time of cooperation.

Many theories have been developed to explain the framing effects, among those there is *Team Reasoning*, proposed by Bacharach (1999, 2018), a model that theorize the individuals within a team pick the strategy profile that is the best outcome for the team. Other models conjectures that the framing activates social preferences that are otherwise absent, it influences the way others may interpret certain behaviors, or it acts as equilibrium selection

[1] We refer to the survey of Dal Bó & Fréchette (2018) for an exhaustive analysis on the determinants of cooperation in a Prisoner's Dilemma.

device[2]. In some way or another, all this models require that a players care about the well being or opinion of others. We, instead, propose a model that works also for self-interested individuals and predicts higher levels of cooperation in the tournament. The main intuition behind this mechanism is that the presence of an opposing team, an enemy, can induce weak links between team members aligning incentives toward cooperation. The player with whom you are playing the PD now become your ally, when it would have been normally your opponent. Moreover we can conjecture that players will attach a non monetary utility to winning the competition. This mechanism is extensively used in work places, sports and personal relationship.

In this study, we want to investigate if it is possible to obtain a better level of cooperation even in the controlled environment of the laboratory, furthermore, we will provide a theoretical justification to our intuition and heuristic observations. The main idea about the theoretical justification is that with the introduction of the tournament, and a related utility from winning it, there is a decrease in thresholds of the discount factors necessary to have cooperation in a sub-game perfect and risk dominant equilibria. Although, we propose an alternative explanation, this work is not designed to prove or falsify existing models. Our primary goal is to document the effect that a tournament has on behavior.

Our study contributes to the literature of framing and of Prisoner's Dilemma in multiple ways. It contributes to the literature on the theory of framing by proposing a model that does not rely on the assumption that players must care about the payoffs or opinions of others. Moreover, the experiment differentiate itself from previous ones. The participants in our experiment play multiple instances of the game, while in previous works they play the game only once[3]. This allow us to study the evolution of cooperation over time, showing that the effectiveness of competitive framing increases as the subjects earn experience. We find that the introduction of the tournament increases considerably cooperation among players, and most surprisingly, this effect increases over time. This suggests that players learn to coordinate during the experiment and tournament setting is a reliable mechanism for achieving a desirable level of cooperation among members of the same group.

The paper is organized as follows: Section 2 will provide theoretical justifications regarding the effect of a competitive framing in a infinitely repeated PD. In Section 3, we describe in detail the experimental design. The results are presented in Section 4 and Section 5 concludes.

[2]Ellingsen et al. (2012) presents the relevant literature in a more structured way and test the different hypothesis with their experiment.

[3]In both Ellingsen et al. (2012) and Engel & Rand (2014) the subjects play a one-shot PD once, while in Lieberman et al. (2004) they play a 7 round PD once.

2 Theory

In this section, we discuss the theory of the Prisoner's Dilemma and how it reflects in the behavior observed in a laboratory. After defining the relevant mathematical objects we look at the existing experimental literature. This process will help us to understand and it will guide our predictions about the behavior of subjects during and experiment. Finally, we present a theoretical model that gives a possible justification to our experimental design. Our model, differing from other already present in the literature, works even for self-interested individuals[4].

2.1 Infinitely Repeated PD and Determinants of Cooperation

Game theory offers us a tool to model the tension between selfish incentives and social efficiency, the prisoner dilemma (PD from now on). PD is 2×2 game in which players can choose between cooperation and defection. Joint cooperation leads to a reward payoff (R), the tension is introduced by a temptation payoff (T) achieved by defecting while the other player cooperates, leaving the other player with the sucker payoff (S). Mutual defection leads to a punishment payoff (P). In order to define a PD it is required that T>R>P>S. Often it is also required that 2R>T+S, this ensures that cooperation generates a higher combined outcome, and therefore, alternating between cooperation and defection is not more profitable than joint cooperation. In order to simplify the exposition, we perform the transformation adopted by Dal Bó and Fréchette (2011), allowing us to define the game using only two parameters, g which is the gain from defection when the other player cooperates and ℓ which is the loss from cooperation when the other player defects.

1/2	**C**	**D**
C	R	S
D	T	P

1/2	**C**	**D**
C	$\frac{R-P}{R-P} = 1$	$\frac{S-P}{R-P} = -\ell$
D	$\frac{T-P}{R-P} = 1 + g$	$\frac{P-P}{R-P} = 0$

Table 1: PD Row Player's Payoffs, Original and Normalized

[4]With self-interested individuals, we mean subjects that care only about themselves, but not necessarily only about their material payoff, which are usually called selfish in the literature. In our model, the players care about their material payoff and the utility they derive from winning a competition, hence they are self-interested, but not selfish.

Since cooperation is a dominated action in the one-shot game, standard game theory tells us that repeated interaction is necessary for a rational and payoff maximizer player in order to have credible punishments and rewards that can lead to cooperation in a subgame perfect equilibrium (SPE).

The first wave of experiments that investigated the role of repetition in PD were conducted by Roth and Murnighan (1978) and Murnighan & Roth (1983). They introduced a random termination rule in which at every stage there is a probability δ that the game will continue to the next round. This probability δ replaces the discount factor and allows the experimenter to implement an infinitely repeated game in the laboratory. These authors found that cooperation increases with δ, but not monotonically. A second wave of experiments confirmed those findings with stronger evidences. Dal Bò (2005) found that cooperation increases fourfold from an one shot PD to a indefinitely repeated PD with $\delta = 0.75$. The effect is stronger also because in that experiment participants played several supergames, implying that learning is also a driver of cooperation. A more detailed meta-analysis is available in Dal Bò & Fréchette (2018). In this paper the authors conclude that *"Cooperation is increasing in the probability of future interactions, and this effect increases with experience"*.

The continuation probability is not the only determinant for cooperation. Since a PD is defined by its parameters g and l, it is logical to assume that also these parameters have an important role. For any payoff matrix we can calculate the minimum δ required to sustain cooperation in a SPE:

$$\delta^{SPE} = \frac{g}{1+g}.$$

In a similar way, we can compute the minimum delta such that cooperation is part of a risk-dominant equilibrium:[5]

$$\delta^{RD} = \frac{g+l}{1+g+l}.$$

In the same aforementioned survey Dal Bò & Fréchette (2018) conclude that on average cooperation is greater in treatments in which it can be supported in equilibrium, and even greater when cooperation is risk dominant, but these factors do not imply that a majority of subjects will cooperate. Moreover has been shown that the signs and the magnitudes of $(\delta - \delta^{SPE})$ and $(\delta - \delta^{RD})$

[5]Harsanyi & Selten (1988) define risk dominance for 2×2 games. It is possible to extend the concept of risk dominance to repeated games using auxiliary 2×2 games that implement specific equilibrium strategies. For more reference see Blonski & Spagnolo (2015).

are statistically significant and predict the amount of cooperation achieved in a treatment. Moreover, in Dal Bò & Fréchette (2018) the authors study also the evolution of cooperation in relation to the game parameters, in this case expressed using δ^{SPE} and δ^{RD}. They find that cooperation decreases with experience when it is not risk dominant but increases with experience when it is risk dominant. We will rely on these findings to justify and predict the effects of our treatment.

Despite of the statistical significance of the indexes $(\delta - \delta^{SPE})$ and $(\delta - \delta^{RD})$, a large amount of variation among treatments remains unexplained, moreover there is a lot of heterogeneity among participants. Therefore, it is obvious to expect that personal traits and preferences could be a driver of cooperation. Those determinants have been deeply studied in the vast literature on the Prisoner's Dilemma, but we decide to do not focus on them and to try to find an explanation that can work even for self-interested individuals.

2.2 Competitive Framing in Infinitely Repeated PD

In this section, we investigate how a competitive framing can affect the equilibria of the game. Theory of framing studies how different labeling of the game or the actions affects the behavior. The possible explanations are numerous and they rely on a multitude of general principles. Ellingsen et al. (2012) summarize and develop the existing theory and we defer to this paper for more detailed analysis. In synthesis, the most prominent theories assume that players have altruistic traits or care about the opinion of others, but they are frame dependent and they are triggered only if the framing of the game suggests so. Alternatively, other theories posit the framing affects the beliefs and not the preferences, and it acts as coordination device for the possible equilibria. Notice that also this class of theories assumes that the players have other regarding preferences, therefore even if the game form is a PD, the actual game resembles a different one. The model proposed here differs from the existing one because we don't assume any type of other regarding preferences or image concerns. We assume that the players are purely self-interested. Although we make less assumption about players' preferences than the previous theories, it must be noted that our model relies on a particular setting, a direct competition among groups of players. From now on we refer to this competition as tournament. Previous theories assumed that very few aspects of the game were changed, usually only the labeling of the game or actions. Our changes to the game are more drastic, but it doesn't means that our results are less applicable to real life situations. Given the fact that re-framing a situation is cost-less, our setting can be efficiently implemented in social context in which a social dilemma can arise, let alone all the situations in which competition is naturally present. After this preamble we proceed to present our model.

The Model In the tournament two pairs, teams from now on, of players play an infinitely repeated PD. Players are informed about the presence of an opposing team. In order to win the competition a pair of players must achieve the highest cumulative sum of aggregate payoffs (Points from now on). Winning the tournament does not give any additional material payoff to the winners, moreover the actions of one team don't have any direct effect on the payoffs of the other team. The payoff matrix is the same for all players.

Let now assume that each player assigns a positive non monetary utility upon winning the tournament, $W \geq 0$, in addition to the monetary payoff of the game. Therefore, a player will receive this extra utility only if the points of his team are higher than the ones achieved by the other team. Moreover, we assume that each player has a conjecture about the cooperation achieved by the other team, [6] we will call it X and it will represents the points believed to be necessary to beat the other team [7]. In the presence of the tournament a player must consider the utility W provided by the winning when he decides his strategy. This translate in two main results.

Proposition 1. *Let W and X be the utility given by winning the tournament and the conjecture about other team's points, respectively, then the minimum discount factor necessary to have cooperation as part of a SPE in the presence of a tournament δ^{SPE^*} is lower than δ^{SPE} in absence of the tournament. Moreover δ^{SPE^*} is equal to:*

$$\delta^{SPE^*} = \frac{g - W(\mathbb{1}_{(\frac{2}{1-\delta}>X)} - \mathbb{1}_{(1+g-l>X)})}{1 + g - W(\mathbb{1}_{(\frac{2}{1-\delta}>X)} - \mathbb{1}_{(1+g-l>X)})} \leq \frac{g}{1+g} = \delta^{SPE}.$$

Proposition 2. *Let W and X be the utility given by winning the tournament and the conjecture about other team's points, respectively, then the minimum discount factor necessary to have cooperation as part of a risk-dominant strategy in the presence of a tournament δ^{RD^*} is lower than δ^{RD} in absence of the tournament. Moreover δ^{RD^*} is equal to:*

$$\delta^{RD^*} = \frac{g + l - W(\mathbb{1}_{(\frac{2}{1-\delta}>X)} - \mathbb{1}_{(0>X)})}{1 + g + l - W(\mathbb{1}_{(\frac{2}{1-\delta}>X)} - \mathbb{1}_{(0>X)})} \leq \frac{g+l}{1+g+l} = \delta^{RD}.$$

[6] We assume for simplicity a degenerate (i.e. Dirac's Delta) belief. Results generalize to any distribution.

[7] We assume that W and X are homogeneous across player for sake of simplicity. Results will hold also with heterogeneous player because we will simply take $\delta^{SPE^*} = \max\{\delta_1^{SPE^*}, \delta_2^{SPE^*}\}$.

To prove the first proposition, we followed the proof of Nash reversion introducing the utility W and taking into account the points generated by each strategy. We followed the same logic to prove proposition 2, while following the proof of Blonski & Spagnolo (2015). The detailed proofs can be found in the appendix A.

In light of these two results and the evidences from previous experimental studies, we can justify our expectation of a positive effect of the treatment if the continuation probability δ above δ^{SPE}, but below δ^{RD}. If the utility from winning the tournament is high enough for the participants, we will observe levels of cooperation and learning patterns that resemble those found in game in which cooperation is a risk-dominant strategy.

3 Experimental Design

In our between-subject experiment, the participants play an indefinitely repeated prisoner's dilemma in a lab-experiment. We implement the design adopted in one of the treatments of Dal Bò & Fréchette(2011), namely one in which cooperation can be sustained in a SPE but not in a risk dominant-strategy. The indefinitely repeated prisoner's dilemma is a multi-stage game in which, at each round, the participants play a PD, choosing between two actions, *Cooperate* or *Defect*[8]. At the end of each round, there is a fixed and known probability δ that the game will continue to the next round and the participants will play again the same game with the same partner. We call supergame the series of consecutive stage games played with the same partner. Each stage game has the following game parameters: Reward payoff (*Cooperate, Cooperate*) 32, Punishment payoff *(Defect, Defect*) 25, Temptation payoff (*Defect, Cooperate*) 50, Sucker payoff (*Cooperate, Defect*) 12. These payoffs are expressed in Experimental Currency Units (ECU). These parameters remain fixed during the whole experiment and they are the same in the every treatment. The game form of each PD game is represented in Table. 2

		The other's choice	
		Cooperate	*Defect*
Your choice	*Cooperate*	Reward payoff (32, 32)	Sucker payoff (12, 50)
	Defect	Temptation payoff (50, 12)	Punishment payoff (25, 25)

Table 2: Game payoffs

At the end of every stage game, each player receive a feedback about the action played by his partner and the corresponding outcome. These information are store until the end of the supergame and they are displayed on screen in a history box that contains the actions played in previous rounds by both players and relative outcomes.

When a supergame ends, new pairs are randomly formed and a new supergame, with the same rules, starts. We set the continuation probability to

[8]During the experiment the actions will be labeled as action 1 and 2 respectively. This is done to avoid unwanted framing effects that may arise when a non neutral labeling of the actions is used.

0.75. The participants have 50 minutes to play as many supergame as possible, and their earning is computed as the sum of the outcome of every stage game. We refer to the subjects that play using this set of rules as *Control* group, or simply *Control*.

Out experimental design consists of a single treatment manipulation. In the treatment group (henceforth *Tournament*) the rules of the game are identical to those in the *Control*, with a single exception: we set up a competition among pairs of players (henceforth *Teams*). The rules of the competition are simple, two teams are matched randomly and the team that achieves the highest cumulative sum of payoffs at the end of the supergame is elected winner. The result of the competition, win, loss or tie, is displayed at the end of each supergame. Beware that winning the competition does not grant any additional monetary payoff, and participants are explicitly informed about that. Moreover the competition is strengthened by using a different language in the instructions; the experiment is explicitly called tournament, supergames are called matches, the pairs are called teams and the person with whom participants play the game is called teammate. See the differences in the instructions in appendix B.

Since the team that wins doesn't earn extra money, the tournament must be considered a framing effect. This is crucial to our investigation. Our goal is to study if it is possible to achieve more cooperation in situations similar to a PD by simply introducing a competition, deflecting the tension toward a common foe. If we let the winner earn more, we won't be able to disentangle the two effects, since it would have been impossible to understand if the subjects cooperated more due to the framing or the economic incentives.

3.1 Issues and Concerns

Infinite Repetition in Laboratory: Clearly, it is impossible to implement a real infinite repetition in a laboratory, therefore we adopt the methodology used in Roth & Murnighan (1978). We introduce a random termination rule, namely at the end of each round, there is a probability δ that the game continues and a probability $1 - \delta$ that the game ends. This probability is known to the players and it remains fixed for the whole duration of the experiment. Under the assumption of risk neutrality this termination rule induces the same preferences over outcomes as if the game were played with infinite repetition. The first issues with this methodology is that players potentially are not able to understand correctly probabilities and how these relate to the expected length of the supergame. Experimental evidences reported in Murninghan & Roth (1983) and Dal bò (2005) shows that participants, although not perfectly, have a good understanding of how δ relates to the average length of a supergame.

In Sabater-Grande & Georgantzis (2002), the authors find that subjects during the experiment react differently to the presence of a continuation probability based on their level of risk aversion. Although acknowledging this fact, we still believe that this doesn't undermine the validity of our design. All the parameters of the game, including the continuation probability, are the same in the *Control* and in the *Tournament*, this means that the effect observed in Sabater-Grande & Georgantzis (2002) is present in both groups. Moreover, the balance test confirms that the two samples don't differ significantly in risk attitudes. Since we are primarily interested in the effect of the treatment variable, and the aforementioned effect is present and equally relevant in both groups, it should not jeopardize our results.

3.2 Experimental Procedure

We recruited 94 participants (46 participants in control and 48 in treatment) from the subjects pool of the university of Côte d'Azur (Nice, France) using ORSEE (Greiner, 2015). The subjects pool includes students from many disciplines. The experiment was programmed using zTree (Fishbacher, 2007) and run at the Laboratoire d'Economie Expérimentale de Nice (LEEN). The payoffs are expressed in Experimental Currency Units (ECU), and, at the end of the experiment, participants are paid 0.005€ for each ECU earned during the experiment. The average payment was 21.42€, including the 5€ show up fee, and the experiments lasted on average 75 minutes. We ran 6 sessions, 3 for the *Control* and 3 for the *Tournament*. Each participant played exclusively in one of the two groups.

The experiments took place between September 23rd and September 24th 2020. At that moment, there were in places rules to ensure the safety of the participants and the experimenters, which were meant to minimize the risk of spreading the virus. Masks were mandatory during the experiment, the work stations were sanitized before and after each session and there was a limit of 16 people inside the laboratory. At the end of the experiment participants filled a brief questionnaire in which they self reported about: socio-demographics, generalized trust, risk aversion, altruism and rivalry. [9]

[9] See the questionnaire in the appendix section B.3.

4 Results

In this section, we are going to test our hypothesis analyzing the data we gathered during the experimental sessions. Firstly, we proceed with a description of the sample and we run a balance test in section 4.1. Then, we test our main hypothesis looking for a treatment effect over cooperative behavior. Or main result is shown in section 4.2. In section 4.3, we study the evolution of subjects' decisions over time, looking for learning processes similar to those found in Dal Bò & Fréchette (2011). Next, in section 4.4 we investigate if reciprocity plays a different role in the decision-making process in the two treatments . In addition, in section 4.5, we investigate if demographics and self-reported personal traits, may influence the subjects' choices. We end in section 4.6 with an exhaustive regression analysis.[10]

4.1 Descriptive Statistics

We begin our preliminary investigation by depicting the descriptive statistics of our variables in Table 3. Afterward, we run a balance test, in order to demonstrate that our results are driven by the treatment effect and they are not due to an unbalanced distribution of relevant variables between control and treatment. Table 4 reports the results of the balance test, which confirm the robustness of our results.

[Table 3: Summary of descriptive statistics.]

[Table 4: Balance test.]

4.2 Testing Treatment Effect

The main objective of our study is to examine whether inducing a competitive frame (without additional economic remuneration) is sufficient to foster cooperative behavior, in a game were parameters are such that *Cooperate* is part of a SPE but it is not a risk-dominant strategy. In order to obtain a preliminary intuition, we present, in Figure 1, an histogram that shows the frequency of each choice in each treatment. We observe that in both treatments *Defect* is the predominant choice, in line with the literature Croson et al. (2005). Nevertheless, in the tournament treatment the cooperation is higher.

[Figure 1: Frequency of choices by treatment.]

[10]All the tables and figures referenced in this section are displayed in the appendix section C.

The average cooperation goes from 30.88% in the Control to 36.17% in the Treatment. The difference between Control and Treatment is statistically significant (p-value=0.000; Two-sample Wilcoxon rank-sum test). Our results suggest that the framing has a positive effect on cooperation, despite the fact that the Tournament offers no monetary incentives to do so. The presence of a common opponent induces the subjects to coordinate in order to beat the opponent team. Since cooperation gives a greater amount of points, the desire to beat the other team translates in a higher cooperation.

R1. *Framing the social dilemma in a competitive environment foster cooperative behavior.*

4.3 Evolution of Cooperation

We are also interested in studying the evolution of cooperation over the time in order to check whether the subjects learn and adjust their behavior. Table 5 shows the percentage of subjects that choose to cooperate in the first round of each repeated game in this treatment, with the repeated games aggregated according to the interaction in which they started. We follow the same procedure adopted by Dal Bò & Fréchette (2011). Figure 2 exhibits the percentage of cooperation on the first round and the aggregate of all periods for each session for both treatments, respectively Tournament (sessions 1, 2 an 3) and Control (sessions 4, 5, and 6). While Table 6 and Figure 3 show the frequencies for the Control and Tournament respectively.

[Figure 2: Percentage of cooperation in the first round and aggregate of each repeated game by session.]

[Table 5: Percentage of cooperation of treatments divided by session.]

[Figure 3: Percentage of cooperation of repeated games period in Control and Treatment.]

[Table 6: Percentage of cooperation by treatment.]

To compare inexperienced versus experienced players, we compare behavior in the first ten interactions with those last interactions 102 to 145.[11] We

[11] We do have data on repeated games that started even later, but because there are slight variations in the total number of interactions across sessions, the sample size is stable only up to interactions 102–145.

can observe that experience in Control leads to more free riding, because the percentage of subjects' choosing to cooperate decreases with experience. The opposite is found in the Tournament Treatment, where the percentage of cooperation increases respect the first range of interactions (1-10). These results are statistically significant (p-value=0.000; Two-sample Wilcoxon rank-sum test).

R2. *Experienced subjects reduce cooperation in the Control, while in the Tournament cooperation is sustained and increases over the time.*

The results for the control group are in line with those found by Dal Bò & Fréchette (2011). If the parameters are such that cooperation is not risk-dominant we observe a decline over time. Instead, with the introduction of the Tournament, we observe a pattern that resemble the one observed by the two authors in games in which cooperation is risk-dominant. It shows that the framing is able to promote a desirable level of cooperation even in situations in which the game's parameters are not favorable enough.

4.4 Strategic behavior and reciprocity

We are also concerned about the outcomes of the stage games and check if they are consistent with the results previously found. Figure 4 shows the frequencies of the payoffs obtained in the Control and Treatment, respectively. The punishment payoff (*Defect, Defect*) is predominant in both treatment, although it is significantly less frequent in the Treatment (p-value=0.000; Two-sample Wilcoxon rank-sum test). Conversely the reward payoff (*Cooperate, Cooperate*) is more frequent in the Treatment, the difference is statistically significant (p-value=0.000; Two-sample Wilcoxon rank-sum test). This is in agreement with the results previously found, cooperation is more frequent in the presence of the tournament. Despite the difference in the frequency of reward and punishment payoffs there is no statistical difference in the frequencies of sucker/temptation payoffs between the two treatments (p-value=0.237; Two-sample Wilcoxon rank-sum test).

[Figure 4: Frequency of outcomes by treatment.]

Now we investigate the percentage of those who keep cooperating even after observing a defection from their partner. In the control 29.57% of the times *Cooperate* is played after observing defection, while it is played 29.11% of the times in the treatment, the difference is not statistically significant (p-value=0.862; Two-sample Wilcoxon rank-sum test). This means that in the tournament subjects are not willing to sacrifice their own payoff in exchange for a higher chance of winning (remember that the sucker payoff gives more points than mutual defection). Given these results we can conclude that the tournament acts as a coordination device that bolster cooperation.

R3. *The Tournament bolster cooperation by acting as coordination device.*

All the previous findings are also supported by the Spearmn's test, presented in Table 7. The positive correlation between present and previous choices is a further proof that subject persist in playing the same action. The positive correlation with the present and past partner's choices also suggest coordination.

[Table 7: Spearmans' ranks correlation coefficient by choices and past choices.]

Time to Choose and Coordination: We are also interested to check whether the time necessary to take a decision can give us some insight on strategic behavior. Given the rules of the experiment, subject can realize that they can maximize their gains by playing as many games as possible in the 50 minutes. Therefore, they could coordinate and play as fast as possible. If in the control, cooperation is not sustainable for long time, the subjects can coordinate in playing *Defect* as fast as possible, maximizing their earning in the whole experiment, even if they earn less in each round. Instead in the Tournament we observed an higher level of cooperation, therefore we could expect that the subject to coordinate in playing *Cooperate* as fast as possible. In this way they will maximize both earnings and team's points. We can observe in Figure 5 that the time of choice in the control is lower than in the treatment.

[Figure 5: Scatter plot of the time decision by choices and treatments.]

Not surprisingly, we observe that subjects significantly spend more time in order to decide in treatment respect control (p-value=0.000; Two-sample Wilcoxon rank-sum test). This is due to the fact that in the treatment there is an additional layer of strategic thinking. When subjects reason before making a choice they have to consider also the effect that their choice will have on the probability of winning the tournament. This, of course, takes more time. In agreement with our previous results, when subjects select *Cooperate*, on average, they spend 2.541 seconds in the control, and 2.916 seconds in the treatment (p-value=0.000; Two-sample Wilcoxon rank-sum test).

While, when subjects chose *Defect*, on average, they spend 2.426 seconds in the control, and 3.345 seconds in the treatment (p-value=0.000; Two-sample Wilcoxon rank-sum test). Also, the differences between the time of choice between *Cooperate* and *Defect* within each treatment are statistically significant. The results shows that the subjects take less time to play *Defect* in the control. Since cooperation is not risk-dominant in our setting, it means that it is a riskier action, and therefore requires more reasoning. While the subjects take less time to play *Cooperate* in the treatment, this is a further confirmation of the fact that the tournament acts a coordination device.

4.5 Personal Traits

At the end of the experiment, subjects replied to a brief questionnaire, along with their demographics, they self reported some personal traits. The traits reported are generalized trust, risk attitude, altruism and rivalry. Here we investigate how these traits are reflected in behaviors.

Generalized trust seems to play an important role. Both questions about generalized trust conclude that subjects who trust more, significantly play *Cooperate* more often (p-value=0.000; Two-sample Wilcoxon rank-sum test). This result is in line with the theoretical literature on the basin of attraction of AD, as presented in Dal Bò & Fréchette (2018). In the same vein, subjects' more risk loving significantly cooperate more (p-value=0.000; Two-sample Wilcoxon rank-sum test).

R4. *The evidence reports that trustful and risk loving subjects' are more cooperative.*

Moreover, we study whether the demographic variables play a role in the decision process between cooperation of defection. We found evidence of woman playing *Defect* more often, while man are more cooperative (p-value=0.000; Two-sample Wilcoxon rank-sum test). Subjects with higher level of formation select more often to *Defect* (p-value=0.000; Two-sample Wilcoxon rank-sum test). Furthermore, our results suggest that subjects that have more experience in lab-experiments are more prone to defect (p-value=0.000; Two-sample Wilcoxon rank-sum test).

We didn't use the data of self reported altruism and rivalry because we found them unreliable. Performing a standard reliability test, Cronbach's α coefficients, those variables did not satisfied the recommended 0.75 cut-off value.

4.6 Regression Analysis

In this section, we run a more accurate analysis by studying the causal relationship between choices and the other independent variables such as treatment, learning over the time, time to decide, generalized trust, risk aversion and socio-demographic variables. In the Table 9 are presented the the marginal effects of a probit model.[12] The regression shows an overwhelming, and consistent, effect of the treatment over cooperation, supporting our aforementioned results. Consistent with the literature, we find that cooperation decreases over time within each supergame. With a more accurate analysis the time to make the decision seems not to play a significant role.

[Table 9: Marginal effects of Probit regressions]

Furthermore, generalized trust and risk attitude influence positively cooperation. As previously stated in *Result 4*. The socio-demographic variables significantly influence subjects' decisions. Older subjects are significantly less cooperative. As well, it is found a gender effect, females play *Defect* more often. Finally, subjects whom have participated more times in a lab experiment are more likely to defect.

[12]In the appendix section C are reported the results from the Probit model in Table 8 and the margins in Table 9.

5 Conclusions

Many studies investigated the framing effects on a Prisoner's Dilemma, documenting how a competitive framing decreases the cooperation among partners. In this work we design an experiment in which the competition is diverted towards an opposing team. The partner, that was seen as an enemy, becomes a friend in virtue of being an ally against different opponents. We achieve this by implementing a tournament, in which the goal is to accumulated more payoff as a team, but there are no economic incentives to do so.

We find that cooperation is significantly higher with the introduction of the tournament, this finding is robust and it persist upon further investigations. The tournament is in essence a framing, since it does not modify the payoffs or the game form in a meaningful way for a selfish and risk-neutral player. Framing effects are well documented in the literature, and experimenters are aware of those when they design an experiment. Having said that, the most surprising of our findings is that the positive effect of the tournament persist, and get stronger, over time. One could argue that this is a demand effect, namely subjects do what it is asked by the experimenter. The results don't support this argument, since the difference between the two treatments become more evident as subjects gain experience. This suggest that introducing competition in a situation in which there is a possibility for free-riding significantly reduces these problems. Moreover its effect don't vanishes over time, instead become stronger, suggesting that competition could be implemented as a long term solution.

Our results are in agreement with some findings of Dal Bó and Fréchette (2011) that studied the amount and evolution of cooperation is various PD with different game's parameters. They found that cooperation is higher in game where the parameters are such that cooperation is risk-dominant. Moreover, they found that cooperation increases with experience when it is risk-dominant, while it decreases over time in all other cases. We observe the same pattern for the control group that uses the same game form employed in Dal Bó and Fréchette (2011) for the treatment where cooperation is part of an SPE but not risk-dominant. Surprisingly, we observe for the tournament the same pattern that the authors found for treatment with risk-dominant cooperation.

Furthermore, we found that some of the personal characteristics of the subjects influence cooperation. Generalized trust is found to have a positive effect on cooperation. It is not unexpected to notice a greater willingness to cooperate among those inclined to trust others. In the same fashion, risk loving subject are more willing to cooperate, because they have less fear to the other subject taking advantage of them by free-riding. Our results suggest to policy-makers that framing competition is efficient strategy to sustain cooperation over time, because it is less costly respect the alternative by offering greater returns.

References

[1] Michael Bacharach. Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics*, 53(2):117–147, 1999.

[2] Michael Bacharach. *Beyond individual choice.* Princeton University Press, 2018.

[3] Matthias Blonski and Giancarlo Spagnolo. Prisoners' other dilemma. *International Journal of Game Theory*, 44(1):61–81, 2015.

[4] Pedro Dal Bó. Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *American Economic Review*, 95:1591–1604, 2005.

[5] Pedro Dal Bó and G.R Fréchette. The evolution od cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, 101:411–429, 2011.

[6] Pedro Dal Bó and Guillaume R Fréchette. On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature*, 56(1):60–114, 2018.

[7] Morton Deutsch. Trust and suspicion. *Journal of Conflict Resolution*, 2(4):265–279, 1958.

[8] J Richard Eiser and Kum-Kum Bhavnani. The effect of situational meaning on the behaviour of subjects in the prisoner's dilemma game. *European Journal of Social Psychology*, 1974.

[9] Tore Ellingsen, Magnus Johannesson, Johanna Mollerstrom, and Sara Munkhammar. Social framing effects: Preferences or beliefs? *Games and Economic Behavior*, 76(1):117–130, 2012.

[10] Christoph Engel and David G Rand. What does "clean" really mean? the implicit framing of decontextualized experiments. *Economics Letters*, 122(3):386–389, 2014.

[11] Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10:171–178, 2007.

[12] Ben Greiner. Subject pool recruitment procedures: organizing experiments with orsee. *Experimental Tools*, 1:114–125, 2015.

[13] John C Harsanyi and Reinhard Selten. A general theory of equilibrium selection in games. *MIT Press Books*, 1, 1988.

[14] Aaron C Kay and Lee Ross. The perceptual push: The interplay of implicit cues and explicit situational construals on behavioral intentions in the prisoner's dilemma. *Journal of Experimental Social Psychology*, 39(6):634–643, 2003.

[15] Varda Liberman, Steven M Samuels, and Lee Ross. The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin*, 30(9):1175–1185, 2004.

[16] J Keith Murnighan and Alvin E Roth. Expecting continued play in prisoner's dilemma games: A test of several models. *Journal of Conflict Resolution*, 27(2):279–300, 1983.

[17] Alvin E Roth and J Keith Murnighan. Equilibrium behavior and repeated play of the prisoner's dilemma. *Journal of Mathematical Psychology*, 17(2):189–198, 1978.

[18] Gerardo Sabater-Grande and Nikolaos Georgantzis. Accounting for risk aversion in repeated prisoners' dilemma games: An experimental test. *Journal of Economic Behavior & Organization*, 48(1):37–50, 2002.

A Proofs

A.1 Proof of Proposition 1

In order to prove proposition 1, we follow the proof of Nash reversion and we add to each strategy the value of winning the tournament W weighted by the subjective probability of winning given a conjecture about the other team's points. Therefore the equation become the following:

$$\sum_{t=0}^{\infty} \delta^t \cdot 1 + W\mathbb{1}_{(\frac{2}{1-\delta}>X)} \geq 1 + g + \sum_{t=1} \delta^t \cdot 0 + W\mathbb{1}_{(1+g-l>X)}$$

where $\frac{2}{1-\delta}$ are the points obtained by cooperating every round and $1 + g - l$ are to point obtained by the first round of defection, and mutual defection onward. Rearranging the formula we obtain:

$$\delta^{SPE^*} = \frac{g - W(\mathbb{1}_{(\frac{2}{1-\delta}>X)} - \mathbb{1}_{(1+g-l>X)})}{1 + g - W(\mathbb{1}_{(\frac{2}{1-\delta}>X)} - \mathbb{1}_{(1+g-l>X)})}.$$

A.2 Proof of Proposition 2

In order to prove proposition 2, we follow Blonski and Spagnolo (2015). To asses when coordination is risk-dominant we focus only on two equilibria in pure actions, the grim trigger strategy (GT), which is the least risky among cooperative equilibria (proof in Blonski and Spagnolo 2015), and always defect (AD). We build an accessory 2×2 game using only these two pure strategy equilibrium points. According to Harsanyi and Selten (1988) risk dominance in 2×2 games can be determined by comparing the Nash-products of the two equilibria, namely the product of both players' disincentives not to behave according to the equilibrium under consideration. We call these disincentives u_i for GT and v_i for AD, and they are defined as:

$$u_i = \sum_{t=0}^{\infty} \delta^t \cdot 1 - (1 + g) - \sum_{t=1} \delta^t \cdot 0 \geq 0$$

$$v_i = \sum_{t=0}^{\infty} \delta^t \cdot 0 - (-l) - \sum_{t=1} \delta^t \cdot 0 \geq 0.$$

The grim trigger strategy is risk dominated by AD iff $v_1 v_2 \geq u_1 u_2$. From these relations we find that the threshold for δ below which GT is risk dominated is the following:

$$\delta^{RD} = \frac{g + l}{1 + g + l}.$$

Similarly to proposition 1, we add the weighted value of winning the tournament, therefore the relations become:

$$u_i = \sum_{t=0}^{\infty} \delta^t \cdot 1 + W\mathbb{1}_{(\frac{2}{1-\delta} > X)} - (1+g) - \sum_{t=1} \delta^t \cdot 0 - W\mathbb{1}_{(1+g-l>X)} \geq 0$$

$$v_i = \sum_{t=0}^{\infty} \delta^t \cdot 0 + W\mathbb{1}_{(0>X)} + l - \sum_{t=1} \delta^t \cdot 0 - W\mathbb{1}_{(1+g-l>X)} \geq 0.$$

Using the the same procedures as before we get,

$$(l+W(\mathbb{1}_{(0>X)} - \mathbb{1}_{(1+g-l>X)}))^2 - (\frac{1}{1-\delta} - (1+g) + W(\mathbb{1}_{(\frac{2}{1-\delta} > X)} - \mathbb{1}_{(1+g-l>X)}))^2 \geq 0$$

rearranging the formula we obtain:

$$\delta^{RD^*} = \frac{g + l - W(\mathbb{1}_{(\frac{2}{1-\delta} > X)} - \mathbb{1}_{(0>X)})}{1 + g + l - W(\mathbb{1}_{(\frac{2}{1-\delta} > X)} - \mathbb{1}_{(0>X)})}.$$

B Instructions

B.1 Control Treatment

Welcome

You are about to participate in a session on decision-making, and you will be paid for your participation with cash vouchers, privately at the end of the session. What you earn depends partly on your decisions, partly on the decisions of others, and partly on chance.

General Instructions

1. In this experiment, you will be asked to make decisions in several rounds. You will be randomly paired with another person for a sequence of rounds. Each sequence of rounds is referred to as a match.
2. The length of a match is randomly determined. After each round, there is a 75% probability that the match will continue for at least another round. This probability is always the same regardless of the round. So, for instance, if you are in round 2, the probability there will be a third round is 75% and if you are in round 9, the probability there will be another round is also 75%.
3. At the beginning of a new match, you will be randomly paired with another person for a new match.
4. The choices and the payoffs (expressed in points) in each round are as follows:

	The other's choice	
Your choice	**1**	**2**
1	(32 , 32)	(12 , 50)
2	(50 , 12)	(25 , 25)

The first entry in each cell represents your payoff, while the second entry represents the payoff of the person you are matched with.
For example, if:

- You select **1** and the other selects **1**, you each make 32.
- You select **1** and the other selects **2**, you make 12 while the other makes 50.

- You select **1** and the other selects **2**, you make 50 while the other makes 12.
- You select **2** and the other selects **2**, you each make 25.

5. At the end of the 50 min, you will be payed 0.005€ (half of euro cent) for every point you scored individually in every round played during the whole experiment.

6. Are there any questions?

B.2 Tournament treatment

All the framing introduced in the instructions of the treatment that do not appear in control is indicated in italics.

Welcome

You are about to participate in a session on a tournament, and you will be paid for your participation with cash vouchers, privately at the end of the session. What you earn depends partly on your decisions, partly on the decisions of others, and partly on chance.

General Instructions

1. In this experiment, you will be asked to make decisions in several rounds. You will be randomly paired with a teammate for a sequence of rounds. Each sequence of rounds is referred to as a match.

2. **During each match your team will compete against one adversary team randomly chosen between the other teams in this experiment. The team that earns more points at the end of the match will be declared winner.**

3. The length of a match is randomly determined. After each round, there is a 75% probability that the match will continue for at least another round. This probability is always the same regardless of the round. So, for instance, if you are in round 2, the probability there will be a third round is 75% and if you are in round 9, the probability there will be another round is also 75%. **The match will end for both teams at the same time.**

4. At the beginning of a new match, you will be randomly paired with another **teammate and you will play against a new adversary team.**

5. The choices and the payoffs (express in points) in each round are as follows:

   The first entry in each cell represents your payoff, while the second entry represents the payoff of your **teammate**. **The sum of your payoff and your teammate's payoff in each round during the whole match will determine your total team's points in the match.**
   For example, if:

	Teammate's choice	
Your choice	1	1
1	(32 , 32)	(12 , 50)
2	(50 , 12)	(25 , 25)

- You select 1 and the **teammate** selects 1, you each make 32. **The team's points in the round will be equal to 64.**
- You select 1 and the **teammate** selects 2, you make 12 while the **teammate** makes 50. *The team's points in the round will be equal to 62.*
- You select 2 and the **teammate** selects 1, you make 50 while the **teammate** makes 12. *The team's points in the round will be equal to 62.*
- You select 2 and the **teammate** selects 2, you each make 25. **The team's points in the round will be equal to 50.**

If the total points of your team are higher than the total points of the adversary team, your team wins the match, otherwise your team loses.

6. At the end of the 50 min you will be payed 0.005€ (half of euro cent) for every point you scored individually in every round played during the whole experiment. ***Notice that you will not earn any additional money for winning a match.***

7. Are there any questions?

B.3 Questionnaire

Socio-Demographics

- How old are you?
- What is your gender? Male Female
- What is your occupation?
  ☐ Student
  ☐ Employee
  ☐ Unemployed
  ☐ Retired
  ☐ Other
- If you are a student, what is your field of study?
  ☐ Economy and management
  ☐ Social Sciences
  ☐ Arts and Humanities
  ☐ Engineering Sciences
  ☐ Medical studies
  ☐ Other
- What is your level of study?
  ☐ Elementary school license
  ☐ Middle school license
  ☐ High school license
  ☐ Bachelor's degree
  ☐ Post-graduate degree
- How much experience have you had with LEEN before?

Psychological questions: First

- Generally, do you have confidence in the majority of the people, otherwise for nothing "it is better not trust them"? ☐ Yes ☐No
- From 0 to 10, how much do you trust people in general, where 0 indicates "better not trust none" and 10 means "better completely trust"?

  0 1 2 3 4 5 6 7 8 9 10

- For scale from 0 to 10, how do you evaluate your behaviour in front of risk: you are person who avoids risk (1) or you love risk (10)?

0 1 2 3 4 5 6 7 8 9 10

Psychological questions: Second

- Feel indifference to others' misfortunes

  ☐Never ☐Almost Never ☐Sometimes ☐Frequently ☐Almost Always ☐ Always

- Try not to do favors for others

  ☐Never ☐Almost Never ☐Sometimes ☐Frequently ☐Almost Always ☐ Always

- Feel sympathy for those who are less fortunate than me

  ☐Never ☐Almost Never ☐Sometimes ☐Frequently ☐Almost Always ☐ Always

- Love to help others

  ☐Never ☐Almost Never ☐Sometimes ☐Frequently ☐Almost Always ☐ Always

- Avoid competitive situations

  ☐Never ☐Almost Never ☐Sometimes ☐Frequently ☐Almost Always ☐ Always

- Feel that winning or losing doesn't matter to me

  ☐Never ☐Almost Never ☐Sometimes ☐Frequently ☐Almost Always ☐ Always

- Drawn to compete with others

  ☐Never ☐Almost Never ☐Sometimes ☐Frequently ☐Almost Always ☐ Always

- Feel that I must win at everything

  ☐Never ☐Almost Never ☐Sometimes ☐Frequently ☐Almost Always ☐ Always

C Tables and Figures

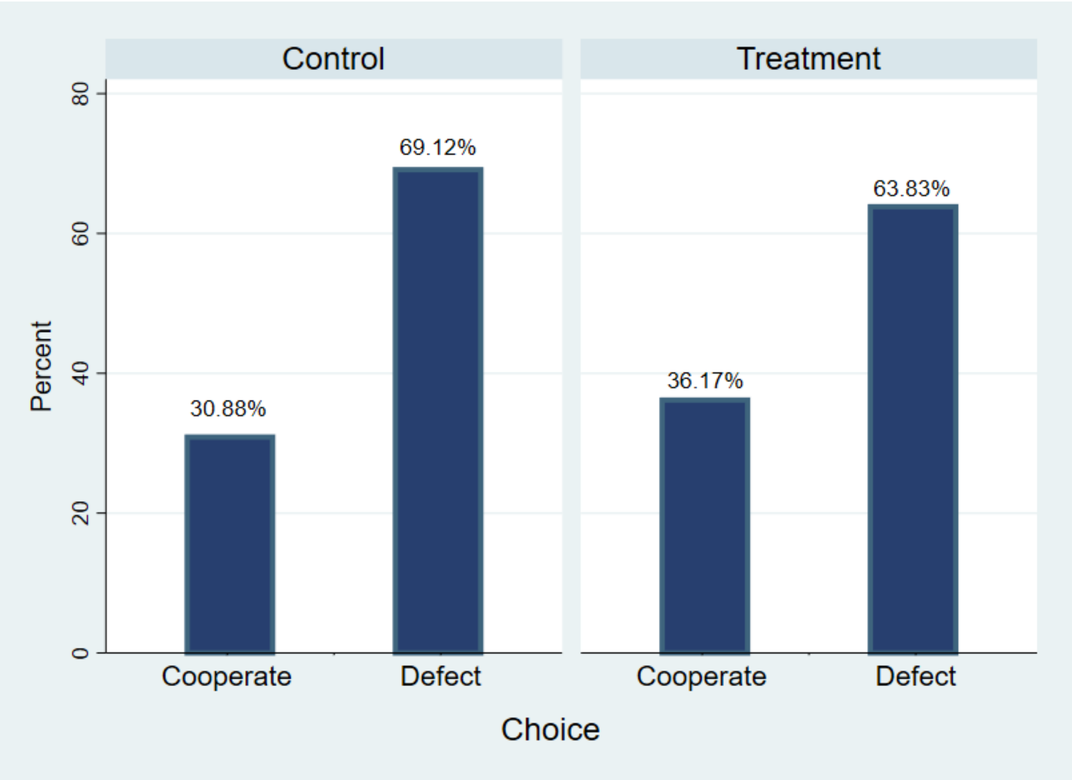

Figure 1: Percentage of cooperation in the control and in the Tournament treatment.

Variables	N	mean	s.d	min	max
Subject	10664	50.030	27.685	1	94
Treatment	10664	0.480	0.499	0	1
Choice	10664	1.666	0.472	1	2
Session	10664	3.632	1.750	1	6
Period	10664	4.250	3.523	1	21
Match	10664	75.084	43.702	1	158
Age	10664	23.977	5.004	18	49
Gender	10664	1.621	0.485	1	2
Occupation	10664	1.282	0.812	1	5
Disciplines	10664	3.406	2.151	1	6
Studies	10664	4.191	0.879	2	6
Experience in lab	10664	3.018	2.764	0	10
Trust (Q1)	10664	0.404	0.491	0	1
Trust (Q2)	10664	5.776	1.804	1	9
Risk loving	10664	5.617	1.881	0	10
Altruism (Q1)	10664	2.594	0.880	1	5
Altruism (Q2)	10664	2.207	0.908	1	5
Altruism (Q3)	10664	4.346	1.246	1	6
Altruism (Q4)	10664	4.524	1.115	1	6
Rivalry (Q1)	10664	3.264	1.183	1	6
Rivalry (Q2)	10664	2.462	1.217	1	6
Rivalry (Q3)	10664	3.287	1.165	1	6
Rivalry (Q4)	10664	3.257	1.451	1	6

Table 3: Summary of the descriptive statistics.

	Control vs Treatment	
Variables	*Z*	*p-value*
Age	1.017	0.309
Gender	1.105	0.269
Occupation	0.672	0.502
Discipline	1.412	0.158
Studies	1.480	0.139
Experience in lab	-0.130	0.897
Trust (Q1)	0.375	0.707
Trust (Q2)	-0.065	0.948
Risk loving	0.686	0.492
Altruism (Q1)	-2.357	0.0184
Altruism (Q2)	0.788	0.431
Altruism (Q3)	-0.371	0.710
Altruism (Q4)	1.182	0.237
Rivalry (Q1)	2.607	0.009
Rivalry (Q2)	0.224	0.822
Rivalry (Q3)	-1.933	0.053
Rivalry (Q4)	-0.543	0.587

Table 4: Balance test.

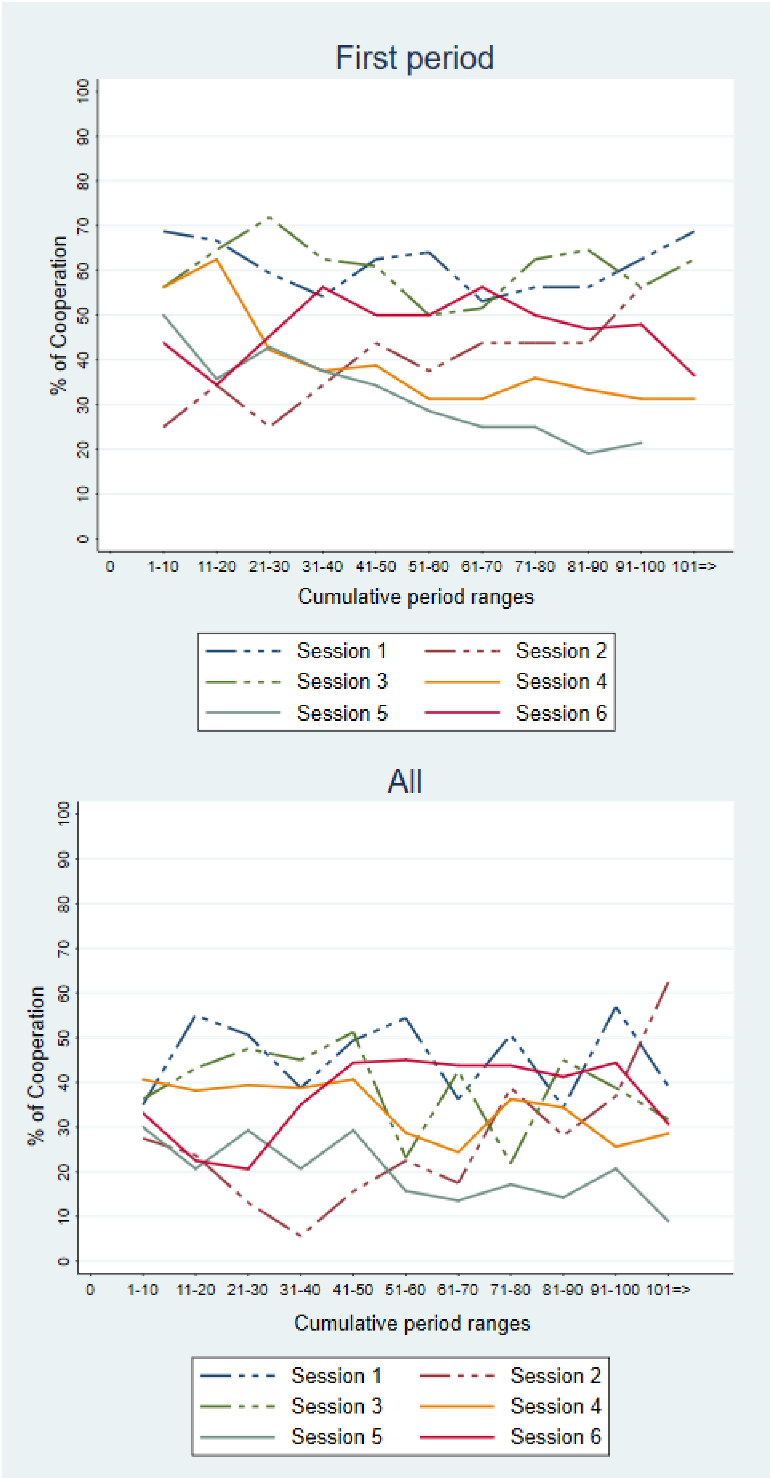

Figure 2: Percentage of cooperation in the first round and aggregate of each repeated game by session.

Repeated games begin in interactions	Control						Treatment					
	Session 4		Session 5		Session 6		Session 1		Session 2		Session 3	
	First period	*All*	*First period*	*All*	*First period*	*All*	*First period*	*All*	*First period*	*All*	*First period*	*All*
1-10	56.25%	40.63%	50.00%	30.00%	43.75%	33.13%	68.75%	35.00%	25.00%	27.50%	56.25%	36.25%
11-20	62.50%	38.13%	35.71%	20.71%	34.38%	22.50%	66.67%	55.00%	34.38%	23.75%	64.58%	43.13%
21-30	42.19%	39.38%	42.86%	29.29%		20.63%	59.38%	50.63%	25.00%	13.13%	71.88%	47.50%
31-40	37.50%	38.75%	37.50%	20.71%	56.25%	35.00%	54.17%	38.75%		5.63%	62.50%	45.00%
41-50	38.75%	40.63%	34.29%	29.29%	50.00%	44.38%	62.50%	49.38%	43.75%	15.63%	60.94%	51.25%
51-60	31.25%	28.75%	28.57%	15.71%	50.00%	45.00%	64.06%	54.38%	37.50%	22.50%	50.00%	23.13%
61-70	31.25%	24.38%	25.00%	13.57%	56.25%	43.75%	53.13%	36.25%	43.75%	17.50%	51.56%	42.50%
71-80	35.94%	36.25%	25.00%	17.14%	50.00%	43.75%	56.25%	50.63%	43.75%	38.75%	62.50%	21.88%
81-90	33.33%	34.38%	19.05%	14.29%	46.88%	41.25%	56.25%	34.38%	43.75%	28.13%	64.58%	45.00%
91-100	31.25%	25.63%	21.43%	20.71%	47.92%	44.38%	62.50%	56.88%	56.25%	36.88%	56.25%	38.75%
101-...	31.25%	28.57%		8.93%	36.46%	30.69%	68.75%	39.06%		62.50%	62.50%	31.70%

Table 5: Percentage of cooperation of treatments divided by session.

Note: The empty cells in the columns of First period appear when the range of aggregate periods do not begin any repeated game (match). Therefore, the range of 10 periods do not contain a first period. The total number of interactions per session are the following: Session 1 is 104; Session 2 is 102; Session 3 is 114; Session 4 is 107; Session 5 is 108; and Session 6 is 145.

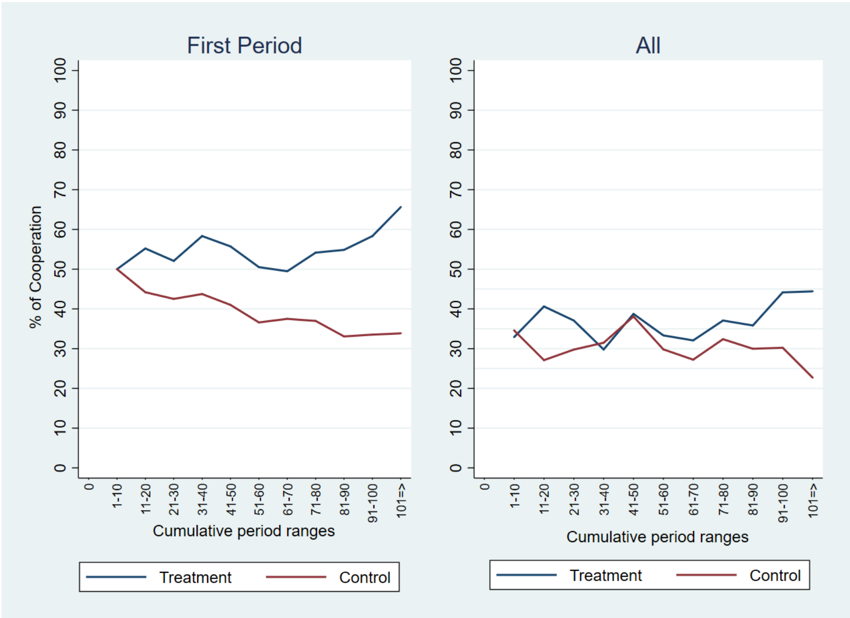

Figure 3: Percentage of cooperation of repeated games period in Control and Treatment.

Repeated games begin in interactions	Control *First period*	Control *All*	Treatment *First period*	Treatment *All*
1-10	50.00%	34.59%	50.00%	32.92%
11-20	44.20%	27.11%	55.21%	40.63%
21-30	42.53%	29.77%	52.09%	37.09%
31-40	43.75%	31.49%	58.34%	29.79%
41-50	41.01%	38.10%	55.73%	38.75%
51-60	36.61%	29.82%	50.52%	33.75%
61-70	37.50%	27.23%	49.48%	32.08%
71-80	39.98%	32.39%	54.17%	37.09%
81-90	33.09%	29.97%	54.86%	35.84%
91-100	33.53%	30.24%	58.33%	44.17%
101-...	33.86%	22.73%	65.63%	44.42%

Table 6: Percentage of cooperation by treatment.

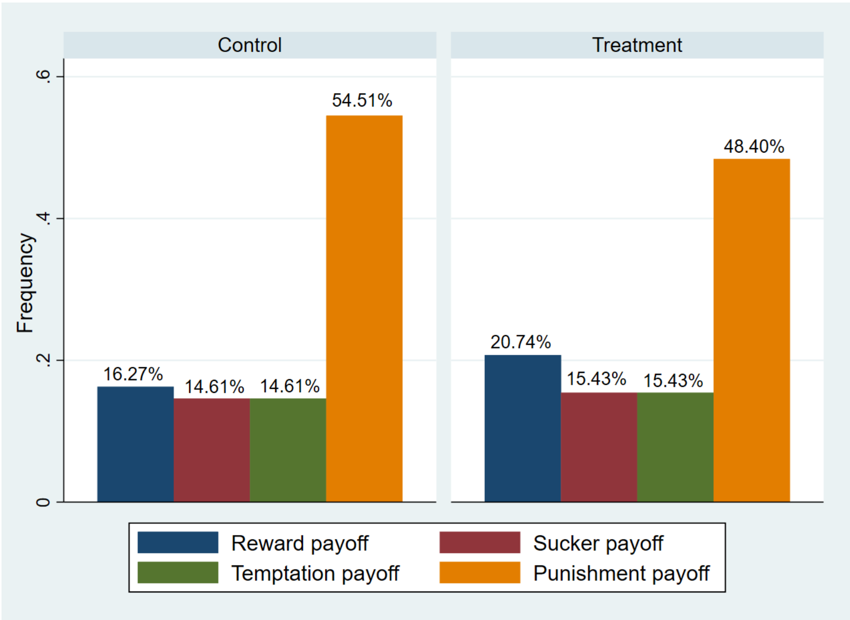

Figure 4: Frequency of the outcomes by treatment.

	Choice
Choice	1.000
Partner choice	0.327***
Past choice	0.438***
Partner past choice	0.436***

Table 7: Spearmans' ranks correlation coefficient between choices and past choices.

Note: Robust standard errors in parentheses ***p<0.01, **p<0.05, *p<0.1. Past choice and Partner past choice variables refer choice of the previous period of own and partner respectively.

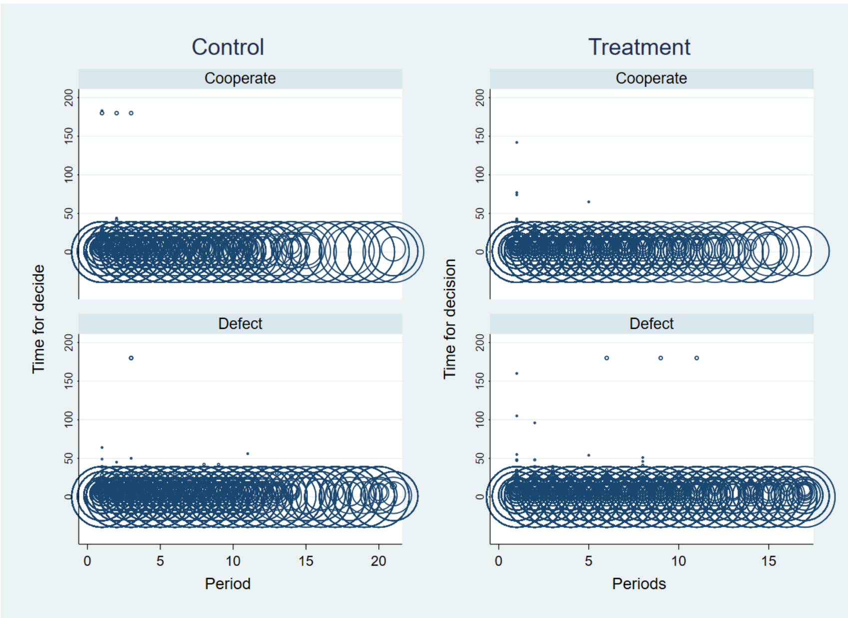

Figure 5: Scatter plot of the time decision weighed: Choices and treatments.

Variable	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Treatment	0.145***	0.175***	0.176***	0.176***	0.120***
	(0.251)	(0.026)	(0.026)	(0.025)	(0.026)
Period		-0.042***	-0.042***		
		(0.005)	(0.004)		
1st Period		0.249***	0.249***		
		(0.034)	(0.034)		
Time to decide			-0.002		
			(0.002)		
Trust (Q1)				0.081***	
				(0.030)	
Trust (Q2)				0.029***	
				(0.008)	
Risk loving				0.080***	
				(0.007)	
Age					-0.009**
					(0.003)
Gender					-0.224***
					(0.027)
Occupation					0.107***
					(0.017)
Discipline					0.018***
					(0.006)
Studies					-0.085***
					(0.015)
Experience in lab					-0.032***
					(0.005)
Constant	-0.499***	-0.401***	-0.396***	-1.173***	0.341**
	(0.018)	(0.029)	(0.030)	(0.055)	(0.100)
Observations	10664	10664	10664	10664	10664
Pseudo R-squared	0.003	0.025	0.025	0.018	0.0173

Table 8: Probit regressions explaining choices.

Variable	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Treatment	0.053***	0.063***	0.063***	0.034***	0.043***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
Period		-0.015***	-0.015***		
		(0.002)	(0.002)		
1st Period		0.090***	0.090***		
		(0.012)	(0.012)		
Time to decide			-0.001		
			(0.000)		
Trust (Q1)				0.029***	
				(0.011)	
Trust (Q2)				0.011***	
				(0.003)	
Risk loving				0.030***	
				(0.003)	
Age					-0.003***
					(0.001)
Gender					-0.081***
					(0.010)
Occupation					0.039***
					(0.006)
Discipline					0.007**
					(0.002)
Studies					-0.031***
					(0.005)
Experience in lab					-0.012***
					(0.002)
Observations	10664	10664	10664	10664	10664

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

Table 9: Marginal effects of Probit regressions.