Strategic Trustworthiness via Non-Strategic Third-Party Reward – An Experiment

Lilia Zhurakhovska[a]

(University of Duisburg-Essen,
Mercator School of Management, Lotharstr. 65, 47057 Duisburg, Germany,
lilia.zhurakhovska@uni-due.de)

Abstract

In modern societies more and more people interact with strangers in one-shot situations. In these situations it might be difficult to trust others. Yet, trust is an essential component of most economic interactions. In this paper (in a one-shot situation) an impartial third party can reward another stranger for being trustworthy towards another unrelated person. By design the reward is costly and cannot be strategically motivated. Subjects strategically increase their trustworthiness towards others if they can anticipate to be rewarded for such behavior by an impartial third party. Impartial third parties reward trustworthiness irrespective of whether it can be anticipated.

Keywords: strong indirect reciprocity, third-party reward, trust game, norms, experiment

JEL: C09, D03, D63

a I gratefully acknowledge financial support from the Max Planck Institute for Research on Collective Goods, Bonn

1. Introduction

> *"Virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence."*
>
> (Kenneth J. Arrow, 1972)

The concept of trust has always played a crucial role in economics (cf. Arrow, 1972). It can be a money-maximizing strategy to trust someone more the higher that person's incentives are to be trustworthy. In repeated settings, one is trustworthy (or cooperates) because one expects others to reciprocate (positively or negatively) one's behavior, and thus mutual trustworthiness leads to higher earnings for the actors. Strategically motivated reciprocity can be carried out by the person who is directly affected by someone's act himself (on direct reciprocity, see, e.g., Andreoni and Miller, 1993; Fehr and Gächter, 2000) or by another person (on indirect reciprocity, see, e.g., Engelmann and Fischbacher, 2009; Rockenbach and Milinski, 2006; Seinen and Schram, 2006).

In modern societies, however, many contacts are anonymous, indirect, and rarely (if ever) repeated. One example of such transactions is anonymous online trades between private persons via platforms such as www.ebayclassifieds.com/ or www.craigslist.org/. Consequently, it is important to study how trustworthiness can be enhanced by the anticipation of a reward (or punishment) by an impartial stranger who does not have an incentive to reciprocate. I.e., in this paper, "positive strong indirect reciprocity"[1] is defined as a third-party reward in situations in which any strategic concerns for the third party can be excluded.[2] The paper provides evidence that positive strong indirect reciprocity exists; it is anticipated by potential recipients; and, most importantly, it can change these recipients' previous behavior in an efficient way.

In the present experimental paper, a trust game is implemented and it is followed by a helping game[3] with a different group composition, i.e., the trustee of the trust game in one group becomes a receiver in a helping game in a different group. The helper has the possibility to reward a recipient conditional on his performance in the trust game. In the Baseline, subjects first play the trust game and receive the instructions for the helping game only afterwards, while in the Anticipation treatment, subjects are informed about the content of both games at

1 According to Gintis (2000, p. 169) "*A strong reciprocator is predisposed to cooperate with others and punish non-cooperators, even when this behavior cannot be justified in terms of self-interest, extended kinship, or reciprocal altruism.*" That is, "strong reciprocity" cannot be explained by any kind of strategic concerns.

2 Carpenter and Matthews (2004) and Carpenter, Matthews, and Ong'ong'a (2004) relate to "*strong indirect reciprocity*" as "*social reciprocity*".

3 The helping game is, in fact, a simple dictator game with an efficiency factor. The game has first been used in Nowak and Sigmund (1998). However, they introduce a repeated setting in which helpers could accumulate image scores over the periods.

the beginning of the experiment. Since the experiment is one-shot, any strategic concerns for the helpers are ruled out in both treatments.

Many helpers make positive transfers to trustees and send significantly more money as the trustees' return transfers increase. Helpers apparently care more about socially desirable behavior than about the motives behind trustees' transfers, i.e., on average, helpers' transfers are the same in both treatments. Most importantly, trustees anticipate helpers' behavior if the helping game is announced, i.e., the absolute level of return transfers as well as the relative level to the investments are higher in the Anticipation treatment as compared to the Baseline. The pattern of the return transfers as well as the fact that the treatment difference is mainly driven by male participants indicate that trustees act in a strategic way in the Anticipation treatment. Investments, on average, do not differ between the Anticipation treatment and the Baseline.

To the best of my knowledge, this is the first paper explicitly to study the effect of anticipated positive strong indirect reciprocity on trustworthy behavior.[4] Further, following the call by Almenberg et al. (2011),[5] it is one of the few papers to study positive strong indirect reciprocity. Note that in terms of welfare, withholding reward is the same as punishing. However, it is not clear that this is true from the motivational point of view. Dohmen, Huffman, Sunde (2009) show that positive and negative reciprocity are different concepts. Analogously, denying positive strong indirect reciprocity does not need to be the mirror concept of exerting negative strong indirect reciprocity. Additionally, in this paper, the impartial stranger has a richer strategy set to condition her[6] rewards on the history of her beneficiary than in previous studies allowing for a more detailed analysis of her choices. Lastly, only one study analyzes crowding out of strong indirect positive reciprocity, if its potential recipient can act strategically (Stanca, Bruni, and Corazzini, 2009).[7] Interestingly, in contrast to their study, the present paper finds no crowding out.

The remainder of the paper is organized as follows: first, an overview of the relevant literature is presented; afterwards, the design of the experiment is explained. Next, hypotheses are explored. The results and the statistical analysis of the data are presented in section five. Section six discusses the results, while in the last section, the conclusions are drawn.

4 In Stanca (2009) and Stanca, Bruni, and Corazzini (2009), the anticipation effect of a reward by an impartial stranger on the behavior of a first-mover (instead of an effect on a second-mover in anticipation of a reward by an impartial third-party) is reported. However, this aspect is not the focus of these papers.

5 Almenberg et al. (2011, p. 75, p. 77) state "*While costly punishment has received the lion's share of attention, costly rewarding also plays an important role in human prosociality.*" They continue: "*[A] sizeable amount of evidence exists for the importance of rewarding in human cooperation. Yet the reward-based analog to third party punishment, where I reward you in an anonymous one-shot interaction because you have cooperated with somebody else, remains largely unexplored.*".

6 Throughout the paper, the female form "she" is used for the third parties (and for investors, i.e., for players A in the experimental design) and the male form "he" for the other players.

7 Stanca, Brunni, and Corazzini (2009, p. 237) extend the concept of "*motivational crowding out*" (Deci and Ryan, 2000) to a mechanism, in which "*the extrinsic motivation of the first mover may crowd-out the intrinsically motivated reciprocating behavior of the second mover.*"

2. Literature

I am aware of only five studies on positive strong indirect reciprocity. In contrast to the present study, none of them studies a change in trustworthiness due to anticipation of a reward by an impartial stranger.

The most closely related papers to the present study are Stanca (2009) and Stanca, Brunni, and Corazzini (2009).[8] In both studies, the return transfer in a one-shot variant of the trust game either comes from the recipient of the investment or from a stranger.[9] In contrast to the present study, both papers focus on strong indirect reciprocity (return transfers) and do not analyze in detail the effect on behavior of a player in anticipation of strong indirect reciprocity (a change in investment). Stanca (2009), however, does report not finding significant differences in the investments. Thus, in his experiment the strategic motives for the players awaiting direct or indirect reciprocity do not seem to matter. Stanca, Brunni, and Corazzini (2009) hypothesize that the motives behind the reciprocated action can crowd out strong indirect reciprocity (rewards by impartial strangers). The treatment difference here is whether the first mover knows that the second stage will follow. The results confirm their hypothesis, i.e., if the strong indirect reciprocator knows that the first mover was aware of the second stage, she transfers a smaller amount compared to a situation in which the second game was not announced. Notably, the results in the present study are not in line with their hypothesis.

Three further studies on reward by an impartial third party are related. Almenberg et al. (2011), Nikiforakis and Mitchell (2014) and Servátka (2009) each implement a one-shot dictator game where a third party can sanction (punish and/or reward) the dictator. Due to the information structure in these studies, behavioral changes in anticipation of reward (or punishment) cannot be studied in these setting. In the studies either in all treatments (Almenberg et al., 2011; Nikiforakis and Mitchell, 2014) or in no treatments (Servátka, 2009), dictators are aware of the presence of the third party and her sanctioning possibilities. In sum, the main findings of these studies are: first, selfish behavior is punished while generous behavior is rewarded, and rewards are at least as common as punishments; second, dictators transfer most if third parties have the choice between punishment and reward; third, dictators' choices have a stronger impact on third parties' actions than social influence and identification of dictators.

Quite a few studies analyze negative strong indirect reciprocity. Only one recent paper examines the anticipation effect of punishment. Balafoutas, Nikiforakis, and Grechenig (2014) demonstrate, in a one-shot, three-player taking game, that taking rates decrease in anticipation of non-strategic punishment by an impartial third party. Furthermore, third-party punishment significantly decreases if counter-punishment directed towards the third party is allowed.

8 Here only the papers on the topic of downstream/social indirect reciprocity (A acts towards B and C acts as a reaction to this in a certain way towards B) are discussed, since these papers are most relevant for the present study. Nevertheless, it is important to mention that there are also interesting papers on generalized/upstream indirect reciprocity (A acts towards B and B acts as a reaction to that in a certain way towards A). Notable examples are Boyd and Richerson (1989) and Greiner and Levati (2005).

9 He calls it a gift-exchange game.

However, the focus of their paper lies on counter-punishment, which could be interpreted as an emotional reaction or revenge. Carpenter and Matthews (2004) run a repeated public-goods experiment. In one treatment, only members of the own group can be punished, while in the other treatment, also members of another group can be punished. The authors find evidence for the existence of negative strong indirect reciprocity, i.e., members of stranger-groups are punished. Fehr and Fischbacher (2004, p. 66) suggest that Carpenter and Mathews (2004) "could not rule out third-party punishment for reasons of self-interest". Fehr and Fischbacher (2004) find strong evidence for third-party punishment in their one-shot, three-person dictator experiment. Bernhard, Fischbacher, and Fehr (2006) run one-shot dictator games with third-party punishment in Papua New Guinea. They find that in-group members are avenged more than out-group members, while the affiliation of the punished person does not play a role for the punishers' decisions. On the contrary, norm violators expect to be punished less if the third party belongs to their peer group than if she belongs to a different group.

3. Design

This section is divided into three subsections. First, the experimental design is explained; second, the motivation behind this design is discussed; and finally, the experimental procedures followed in the experiment are reported.

3.1. Game and Treatments

The game consists of two parts and subjects were aware of that.[10] In the Baseline, subjects receive specific information about the content of each part only immediately before playing the relevant part of the experiment. In the Anticipation treatment, instructions for both parts are handed out at the beginning of the experiment.[11] Subjects are explicitly told that they cannot lose the money they have earned in a previous part in any of the subsequent parts. In the experiment, the experimental currency unit (ECU) is used. All instructions are read out aloud by the experimenter to achieve common knowledge about the procedure.[12]

At the beginning of the experiment, each subject is randomly assigned one of the three roles A, B, or C. Players keep their roles for the two parts of the experiment. The roles A and B are assigned to 11 subjects each and the role C is assigned to 2 players per session. The distribution of the roles is not made explicit to the subjects. They only know which roles there are, their own role, and, at the relevant point in time, the role of their group members. The group composition differs between the parts, i.e., players from part 1 do not meet in part 2. The game is played only once.

10 In total, there are three parts. Part 1 and part 2 are described in this section. Part 3 is a standard risk aversion elicitation experiment (Holt and Laury, 2002). The results of the risk aversion measure are not reported, as these are not relevant for the present study.

11 A treatment that comprises only the trust game was run as well. The results of the *trust game-only* treatment do not differ significantly from the results of the trust game in the *Baseline*; therefore, they are not reported here.

12 An English translation of the instructions is provided in the Appendix B.

Part 1:

In part 1, the reduced trust game (TG) (first introduced by Berg, Dickhaut, and McCabe, 1995) is played by two players, A and B, who move sequentially. The players are endowed with $E^{TG} = 100$ ECU each. At first, player A (from now on called "investor") decides how many ECU she wants to send to B (called "trustee" from now on). Her transfer $t^{TG}_{investor} = X \epsilon \{0, 10, 20, 30, 40, 50, 60\}$ (from now on called "investment").

The investment is tripled by the experimenter and then transferred to the trustee. In case of a positive investment, the trustee can make a return transfer to the investor. His return transfer is $t^{TG}_{trustee} = X * Y$, where $Y \epsilon \{0, 1, 2, 3\}$. Trustee's decisions are elicited via the strategy method (Selten, 1967), i.e., Trustees decides about Y for each possible X.[13] After the trustee has made his decision, all players are informed about their payoffs from the trust game (π^{TG}_{i}). These are:

$$\pi^{TG}_{investor} = E^{TG} - t^{TG}_{investor} + t^{TG}_{trustee} \text{ for the investor}$$

$$\text{and } \pi^{TG}_{trustee} = E^{TG} + 3 * t^{TG}_{investor} - t^{TG}_{trustee} \text{ for the trustee.}$$

Players C (from now on called "outsider") do not take any decision in the trust game, nor are they informed about the decisions of the other players. Outsiders receive a fixed payoff of $\pi^{TG}_{outsider} = 100$.[14]

Part 2:

For the helping game (HG) (similar to the dictator game: Forsythe et al., 1994), played in part 2, new groups with two players are formed. Each group consists of a "helper" and a "recipient". The randomly selected recipient is either a former trustee or a former outsider, while the helper previously had the role of an investor. The participants of a group in the helping game have not been matched in a group in the trust game, i.e., absolute stranger matching is implemented. A helper is endowed with $E^{HG} = 100$ ECU. She can transfer any natural number of ECU (t^{HG}_{helper}) from 0 to 100 to the recipient.

The transfer is tripled by the experimenter and then transferred to the relevant recipient. Helper's decisions are elicited using the strategy method (Selten, 1967).[15] She has to make a decision for every possible composition, i.e., she has to state how much she wants to transfer in case the recipient is former trustee or a former outsider. Additionally, supposing a former

---

13 Figure 1A in the Appendix displays player B's strategy table. Brandts and Charness (2011, p. 392) show that the main limitation of the strategy method is that it "provides a lower bound for testing for treatment effects." A similar argument is made in Fischbacher, Gächter, and Quercia (2012).

14 Note that, in case of zero investment, $\pi^{TG}_{investor} = \pi^{TG}_{trustee} = \pi^{TG}_{outsider} = 100$.

15 Again, one could argue that the strategy method prompts subjects to take different decisions for different situations. However, the results show that helpers indeed condition their transfers on the relative return transfers of the trustees, but less on the investments that trustees receive. Furthermore, helpers' transfers are more strongly correlated with trustees' history in the *Anticipation* treatment compared to the *Baseline*, which again cannot be explained by the use of the strategy method.

trustee is the recipient, she can make her decision conditional on his history in the TG, i.e., she can make her transfer conditional on any possible return transfer $t_{trustee}^{TG}$ of the former trustee to his investor given any possible prior investment $t_{investor}^{TG}$ he might have received. This means that the helper makes 26 transfer decisions.[16] Only the payoff-relevant transfer decision is realized, based on whether the recipient is in fact a former trustee or a former outsider and, if applicable, based on the former trustee's history. After the decisions have been made, the players are informed about the actual group composition and the relevant transfer.[17] The payoffs in the helping game (π_i^{HG}) are:

$\pi_{helper}^{TG} = E^{HG} - t_{helper}^{HG}$ for the helper and $\pi_{recipient}^{TG} = 3 * t_{helper}^{HG}$ for the recipient.

After the subjects have finished the experiment, they complete questionnaires regarding their attitudes towards trust, risk, and reciprocity, as well as demographics.

3.2. Motivation for the Design

A restricted version of the trust game is used for reasons of simplification, i.e., this enables helpers to have a complete overview of every possible situation they might face. This method allows for testing for strong indirect reciprocity of helpers, i.e., it provides detailed data for helpers' transfers conditional on previous behavior of their recipients.

The inclusion of an outsiders, who does not take any actions, helps to identify an individual benchmark for the helpers' general willingness to help. Helpers' transfers to outsider cannot be interpreted as a reward for any previous action. These players have the same income as trustees, who are passive, i.e., who do not receive an investment and who therefore cannot make a return transfer.

In the experiment, an investor becomes the helper in the helping game. One potential critique of the present study could be that the helping game involves a mixture of self-interested and disinterested preferences (Charness and Rabin, 2002). Giving each player feedback about the payoffs after each game and controlling for these payoffs in a regression analysis (instead of trying to elicit beliefs about their earnings) helps to disentangle these motives. Furthermore, participation in the first part of the experiment facilitates the understanding of the strategy method in the helping game.

16 Figure 2A in the Appendix gives an overview of helpers' strategy space.

17 Subjects are informed in the instructions that the players who are not randomly selected to become helpers' recipients, can earn money m in a "small task". It is not made explicit in the instructions how many subjects have to perform that task, nor how much money can be earned in that task. Only the players who are chosen to perform that task receive additional instructions for the task on their computer screen. The task is to count the number of zeros in tables that consisted of 150 randomly ordered zeros and ones (similar to the task used in Abeler et al., 2011). Each correct answer increases m by 50 ECU. The payoff for these players is therefore $\pi_{other\ player}^{TG} = \sum m$. In each session, 10 of the helpers were matched with one former trustee each and one helper was matched with one former outsider. Therefore, one former trustee and one former outsider were not chosen to become a recipient of the helper.

3.3. Experimental procedure

Experiments were run at the University of Bonn. The experiment was programmed and conducted using z-Tree (Fischbacher, 2007). Four sessions with 96 subjects were held, leading to 22 independent observations per treatment for the roles A and B, respectively. Subjects were invited from the University of Bonn using ORSEE software (Greiner, 2015) and had neither participated in trust games nor in dictator or helping games in the past. Most subjects were students. 19 subjects were economics students, 20 were law students. The remaining subjects came from various different disciplines. 44 participants (46%) were female. The average earning (including all games and a show-up fee of 4 Euro) was 12.52 Euro.[18] The sessions lasted 70 minutes on average.

4. Hypotheses

The focus of the present paper is the effect of anticipated strong indirect reciprocity on trustworthiness. Anticipation is, in essence, the same as backward indication. First, hypothesis about helpers' choices need to be made. Then, trustees' reactions to these choices can be predicted.

In the helping game, under the assumption of pure payoff maximization, the theoretical prediction is zero transfers. A self-interested, payoff-maximizing trustee has no monetary incentive for a positive return transfer, irrelevant of the information about the helping game. A rational, self-interested, payoff-maximizing investor anticipates this and does not invest any points in any treatment. From the theoretical point of view, under the assumption of pure payoff maximization and common knowledge, the unique Nash Equilibrium predicts zero transfers in all games and in all treatments.

However, it has been shown that, in helping games, positive transfers are observed (cf., e.g., Forsythe et al., 1994), and in trust games positive transfers are made in both directions (see, e.g., Berg, Dickhaut, and McCabe, 1995). The first is explained by social preferences such as warm glow (Andreoni, 1990). The latter is explained by strong direct reciprocity (modeled by, e.g., Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006; Levine, 1998; or Rabin, 1993). These models assume that an actor has a reciprocity parameter and that the person has a positive utility from punishing unkind action and rewarding kind action. Which actions are perceived as kind depends on the particular model. In fact, the intuition of the model by Levine (1998) is not restricted to two-player direct interactions. In the following, this model will be used to derive (most of the) behavioral hypotheses. In Levine (1998) a player $i = 1, \dots, n$ receives a *direct utility* of u_i and has a coefficient of altruism $-1 < a_i < 1$. Therefore, he receives an *adjusted utility* of $v_i = u_i + \sum_{j \neq i} \frac{a_i + \lambda a_j}{1+\lambda} u_j$, where $0 \leq \lambda \leq 1$. In other words, when $\lambda > 0$, a person derives a positive utility from his own direct utility (which can

18 That corresponded approximately to $15.25 during the first wave of the experiment and approximately to $15.96 during the second wave of the experiment.

be his payoff) and, in addition, a positive utility from rewarding another person for his altruism, i.e., $i's$ utility from $j's$ utility is greater the greater $j's$ coefficient of altruism is.[19] In the model, a player i maximizes her utility given her preferences and her beliefs about the preferences of her recipients. In the present experiment, the helper does not need to form her own beliefs about the preferences of her recipient, since she can infer them from the trustee's return transfers to his investor. Assuming that at least some helpers have a coefficient of altruism $a_i > 0$ and $\lambda > 0$, it follows:

> **Hypothesis 1a:** The higher helpers' recipients' return transfers in the trust game are, the more help will be transferred.

The second key research question is whether helpers' evaluation of trustees' altruism differs between the *Baseline* and the *Anticipation* treatment. Levine's model is a signaling model. A credible signal of trust should be rewarded. In the *Anticipation* treatment high return transfers can be a signal of being a trustworthy type, however it can also be a signal of a strategic action in anticipation of reward. Therefore, from Levine (1998) it follows that helpers' transfers should be higher in the *Baseline* compared to the *Anticipation* treatment. In line with this argument, Falk, Fehr, and Fischbacher (2008) show that intentions matter.[20] If a trustee anticipates that he will be rewarded by a helper in the *Anticipation* treatment, and therefore increases his return transfer for strategic reasons, his act might be perceived as less kind. Falk, Fehr, and Fischbacher (2008) argue that players acting out of intrinsic motives will be rewarded more than other players. This finding suggests that, in the present experiment, helpers' transfers should be higher in the *Baseline* than in the *Anticipation* treatment because, in the *Baseline*, trustees' return transfers cannot be strategically motivated. Similarly, Stanca, Brunni, and Corazzini (2009) find that more strong indirect reciprocity is displayed when strategic motivations can be ruled out. This leads to:

> **Hypothesis 1b:** For every possible return transfer of trustees, helpers' transfers are lower in the *Anticipation* treatment than in the *Baseline*.

Following the model in Levine (1998), one can derive the following prediction about trustees' return transfers in the *Baseline*:

> **Hypothesis 2a:** The trustees in the trust game make higher positive return transfers the higher the investments they receive.

In the *Anticipation* treatment, "*predictive power of the theory is about what we would expect from a signaling model*" (Levine, 1998, p. 605). Since the trustee not only cares about being altruistic (reciprocal) to others (to the investor), but also cares about his personal utility (his payoffs), he has to form beliefs about whether signaling that he is altruistic (trustworthy) can

19 For some parameters the model can also make predictions on negative reciprocity and on unconditional altruism. However, since punishment is not possible in the present experiment, these predictions will not be discussed in this paper. Transfers to outsiders can be explained by helpers' unconditional altruism.

20 Some theories on reciprocity incorporate intentions (e.g., Charness and Rabin, 2002; Rabin, 1993).

be beneficial to him. In the *Baseline*, the expected probability of being rewarded for revealed altruism (trustworthiness) is zero.[21] In the *Anticipation* treatment, however, the trustee knows that a helper has the opportunity to reward him. If at least some trustees expect some helpers to have a coefficient of altruism $a_i > 0$ and $\lambda > 0$, they should expect to receive higher transfers in the helping game the more trustworthy they are, i.e., they anticipate the prediction in Hypothesis 1a.[22] If a positive number of trustees expects their helper's a_i to be large enough to offset the monetary loss from higher return transfers, these trustees have an incentive to make higher return transfers in the *Anticipation* treatment than in the *Baseline*. Consequently, one can derive the following hypothesis:

> **Hypothesis 2b:** The trustees in the trust game make higher positive return transfers in the *Anticipation* treatment than in the *Baseline*.

Costa-Gomes, Huck, and Weizsäcker (2014) find a positive correlation between an investors' investment and his optimism about a return transfer. Hence, if at least some investors anticipate the predictions about return transfers, which follows from Levine (1998), this leads to:

> **Hypothesis 3:** Investors make higher investments in the *Anticipation* treatment than in the *Baseline*.

5. Results

This section presents the experimental data. It is organized as follows: first, the hypotheses concerning the transfers are tested non-parametrically[23] and then robustness of the results with parametric tests including further control variables is checked. The results are presented in the same order as the respective hypotheses.[24] Note that the main research question of the paper concerns trustees' behavioral change in anticipation of helpers' transfers. Thus, the most important results are presented in section 5.2.

5.1. Helpers' Transfers

This section first investigates whether helpers' transfers depend on relative return transfers made by trustees in the trust game and on the investment received by the trustees. Then, treatment differences on helpers' transfers are tested. In addition, it is analyzed which level of the return transfers is particularly strongly rewarded, i.e., what level of return transfers is regarded as especially altruistic and therefore worth to be rewarded.

---

21 One could argue that by not announcing the helping game in the *Baseline*, but by informing the players that a second game will follow, some subjects might expect their actions in the trust game to have an influence of their later payoffs. In that sense, informing subjects about the existence of second game is a more conservative test.

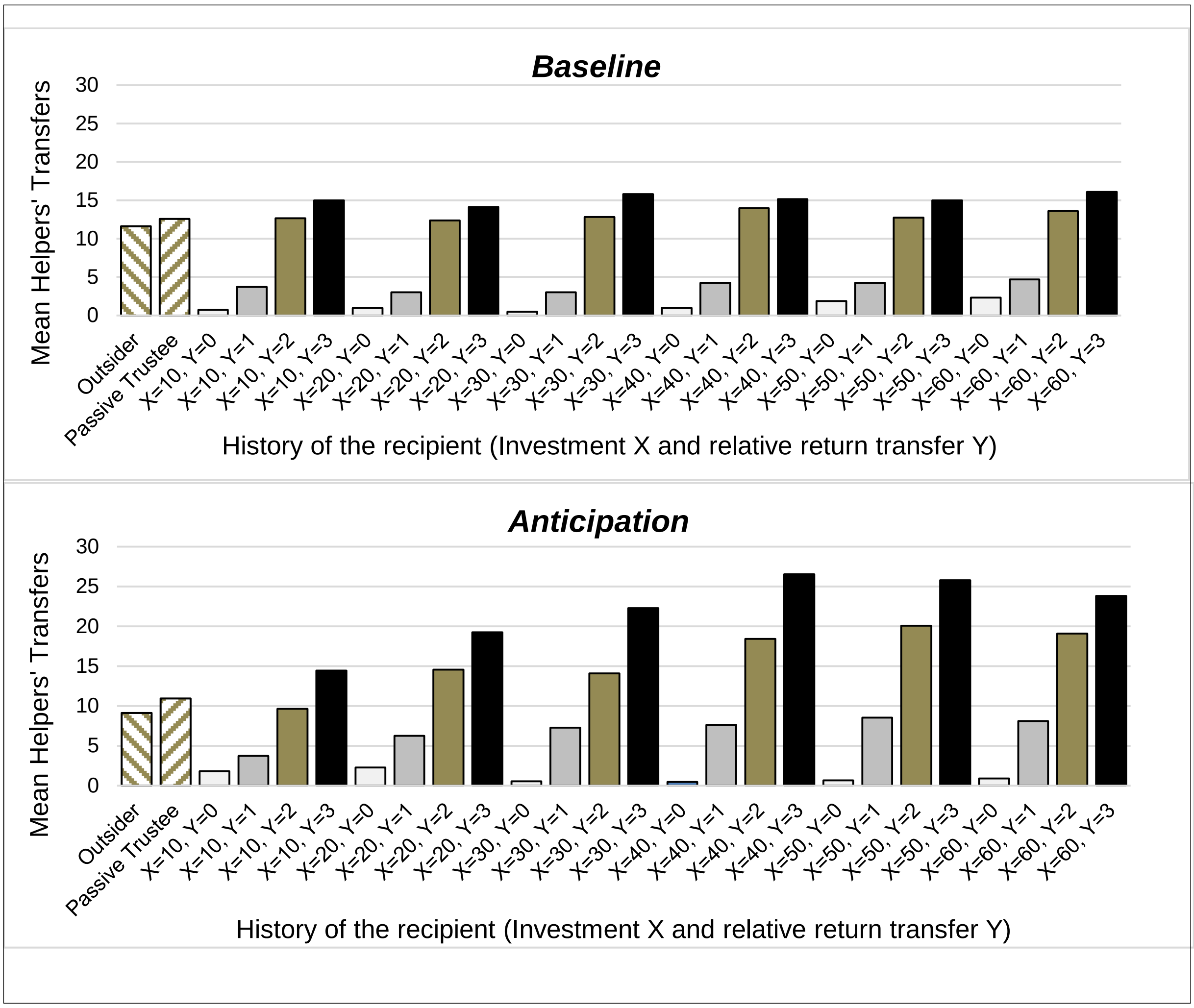
22 If they anticipate as well Hypothesis 1a as 1b, they might be less optimistic about helpers' transfers. Nonetheless, as long as they do not expect helpers to reward *only* intentions, they should expect higher transfers the more trustworthy they are.

23 Throughout this paper, reported p-values are always two-sided, unless stated otherwise.

24 Table 1A gives an overview of all descriptive statistics from all games.

As one can see in Figure 1, higher relative return transfers are rewarded by higher helpers' transfers (Spearman's Rho: $r_S = 0.48$, $p = 0.000$; *Baseline*: $r_S = 0.44$, $p = 0.000$; *Anticipation*: $r_S = 0.53$, $p = 0.000$). Figure 1 furthermore suggests that helpers' transfers for a particular return transfer do not depend on investments (Spearman's Rho: $r_S = 0.02$, $p = 0.621$; *Baseline*: $r_S = 0.01$, $p = 0.788$; Anticipation: $r_S = 0.02$, $p = 0.612$). Overall, transfers are very similar in the *Baseline* and in the *Anticipation* treatment. The simple non-parametric analysis of all average transfers of the helpers yields no significant treatment difference (Mann-Whitney rank-sum $|z| = 1.11$, $p = 0.266$).[25]

Figure 1
Helpers' Transfers by Treatment



The upper figure displays helpers' transfers in the *Baseline* and the lower figure in the *Anticipation* treatment. On the x-axis, the exact condition is displayed, i.e., one can see if the recipient is either outsider or a passive trustee (no investment and thus no opportunity for a return transfer) or an active trustee who has received an investment of X and has made a relative return transfer of Y (Y=0: trustees keeps full transfer; Y=1: trustee returns transfer and keeps rent; Y=2: equal split; Y=3 full return). On the y-axis, mean helpers' transfers are displayed for the particular situation. Standard errors are indicated.

25 No pairwise comparison for a given level of investment and return transfer turn out insignificant (Mann-Whitney rank-sum $|z| = 1.20$, $p \geq 0.229$).

Table 1
Explaining helpers' transfers – comparison *Baseline* and *Anticipation* treatment

Random effects Tobit regressions ("helpers" as group)				
Dependent variable: Helpers' transfers in the helping game to active trustees				
	Model 1	Model 2	Model 3	Model 4
Anticipation	10.05	5.66	9.87	-1.24
	(10.59)	(10.38)	(8.52)	(9.43)
Relative Return Transfers	16.43***	16.43***	16.41***	14.82***
	(.72)	(.72)	(.72)	(1.00)
Investment	.12***	.12***	.12***	.044
	(.04)	(.04)	(.04)	(.06)
*Anticipation**Relative Return Transfer				2.94**
				(1.37)
*Anticipation**Investment				.14*
				(.08)
Own Investment		.03	-.23	-.23
		(.26)	(.23)	(.26)
Own Profit in TG		.36**	.12	.12
		(.17)	(.15)	(.15)
Transfer to Passive Trustee			.34	.71
			(.47)	(.45)
Transfer to Outsider			.71	.32
			(.45)	(.47)
Constant	-47.22***	-82.30***	-62.71***	-56.65***
	(7.99)	(20.41)	(17.12)	(17.25)
N	1056	1056	1056	1056
N of group	44	44	44	44
P model	<.001	<.001	<.001	<.001
Wald Chi2	528.48	530.67	539.19	550.51

Random effects Tobit regressions. Standard errors are presented in parentheses. Observations are clustered on individual level. The *Anticipation* dummy equals 1 for all observations of the *Anticipation* treatment, *relative return transfers* controls for the relative return of a trustee (Y) for a given investment, *investment* controls for the investment the trustee has received, *own investment* is the investment the helper has transfers in the trust game himself to her trustee, *Anticipation*relative return transfer* and *Anticipation*investment* are interaction terms, *own profit in TG* controls for the helper's profit from part 1 of the experiment. *Transfer to passive trustee* and *Transfer to outsider* are the levels transferred to passive players. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively. Left-censored = 577; right-censored = 14.

Random effects Tobit regression models confirm the visual impression.[26] The dependent variable is the helpers' transfers in the helping game. In no model in Table 1 is the dummy variable for the *Anticipation* treatment significant. Model 1 in Table 1 shows that the main determining factor for helpers' transfers is the relative return transfer of trustees, i.e., higher relative return transfers yield higher helpers' transfers.[27] The investment received by the trustees (resulting in a higher or lower absolute return transfer) has a significant, but very small, positive influence on helpers' transfers. The coefficient is not significant once one controls for an interaction between the treatment and the effect of the investment (see Models 4).

Model 2 controls for helpers' own experience in the trust game. It shows that pure willingness to send positive transfers does not lead to higher helpers' transfers (variable *own investment*). However, there could be a small wealth effect: the more a helper has earned in the previous trust game, the more willing she is to help in the helping game.[28] The significance vanishes once one controls for helpers' transfers to passive players (see Model 3), which can be used as a measure of a helpers' individual benchmark of altruism.

In Model 4, an interaction term between the *Anticipation* treatment and the *investment* to the trustee and an interaction term between the *Anticipation* treatment and the *relative return transfer* of the trustee are added. The coefficient of the investment is not significant in this model. On the contrary, coefficients of both interaction terms are significant and positive. The most important result derived from Model 4 in the Tobit regressions is that trustees are treated somewhat differently in the *Anticipation* treatment than in the *Baseline.* In contrast to the prediction in Hypothesis 1b, they receive, on average, for a given history (a particular investment followed by a particular return transfer) a higher transfer from a helper, if they can invest in their good reputation knowing that the helping game will follow. This result contradicts the findings in Stanca, Brunni, and Corazzini (2009). It also calls for a new model of (positive strong indirect) reciprocity in which other factors besides intentions are considered.[29] In the present experiment, helpers seem to care more about socially desirable behavior of trustees than about intrinsically motivated intentions behind the return transfers.[30] An alternative ex-

26 The Tobit regression is used because, in the helping game, helpers' transfers are exogenously restricted with an upper and a lower bound; the lower bound is usually zero-giving. Bardsley (2008) shows that subjects also take money if they have the opportunity in similar situations. In the present setting, this seems plausible, since, as stated before, relative return transfers are rewarded by helpers, while very low relative return transfers lead to very low helpers' transfers and often to transfers of zero. Moreover, as there are 24 transfer decisions per individual (due to the strategy method – only taking into account transfers to active players), random effects models which take individual specific effects into account are in order. The coefficient of the treatment dummy is directly interpretable in the sense that it gives exactly the value of the average marginal effect of the independent variable.

27 Making use of the post-experimental questionnaire data I find that negative reciprocity is negatively correlated with helpers' transfers (Spearman's Rho: rS,= -0.32; p = 0.032). Neither positive reciprocity, nor the gender nor any other personality trait is significantly correlated with helpers' transfers.

28 One could also interpret that coefficient as a proxy for generalized/upstream indirect reciprocity – the more a helper has received in return from her own trustee, the more does she help in the helping game. With the help of the regression, one can disentangle that motive from the social/downstream indirect reciprocity motive of the helper, which is captured in the variable *relative return transfers*.

29 Obviously, other models explaining strong (indirect) reciprocity exist (e.g., Cox, Friedman and Sadiraj, 2008 operate on agents' opportunity sets and discusses the importance of self-serving generosity). Yet, I am not aware of a model that would explain why the data in the present study do not support Hypothesis 1b.

30 One should be cautious in interpreting this result, since in the present experiment helpers arise from the population of investors. Therefore, they might feel belonging to the group of investors. As a result, they might care more about the total earnings resulting from high return transfers of investors than about potential strategic motives of trustees.

planation could be that helpers are not able to distinguish selfishly (or extrinsically) motivated acts from truly pro-social (or intrinsically motivated) acts.

In the regressions, there is strong left-censoring, which indicates that helpers would possibly like not just to give less money to, but even to take money from greedy trustees.

The results from the random effects Tobit regression as well as the non-parametric analysis mainly support Hypothesis 1a, while there is no support of Hypothesis 1b. This leads to:

> **Result 1a:** The higher the relative return transfers a trustee makes in the trust game, the more he receives from a helper. Generally, the investment that a trustee has received does not seem to determine a helper's transfer to the trustee.

> **Result 1b:** Helpers' transfers are, on average, not lower if the helping game is announced. The transfers are more positively correlated with relative as well as with absolute return transfers, if the helping game is announced.

The strategy of helpers includes passive players (outsiders and trustees, who do not receive an investment and can therefore not make any return transfers). Helpers' transfers to passive players can be regarded as a benchmark of how much a helper is willing to transfer to a player who does not have a history.[31] Furthermore, it can be identified which distributive norms achieved in the trust game helpers reward (higher transfer than to passive players) and which they punish (lower transfer than to passive players). A relative return transfer of Y=3 leaves the trustee with his endowment only and reciprocates the investment completely; a relative return transfer of Y=2 leads to an equal split between the trustee and the investor; by a relative return transfer of Y=1, the investor is compensated for his investment and earns as much as his endowment would have been without an investment, while the trustee keeps the complete rent from the investment; the least generous possible relative return transfer in the experiment is Y=0, i.e., this makes the investor worse off than if she had not made an investment and leaves the trustee with the highest possible earning. Wilcoxon signed-rank tests shows that the outsiders are not treated differently than passive trustees ($|z| = 0.65$, $p = 0.514$; *Baseline* only: $|z| = 0.90$, $p = 0.369$; *Anticipation* only $|z| = 0.20$, $p = 0.844$). For almost any investment, for a relative return transfer of Y<2 a trustee is "punished"[32] (Wilcoxon signed-rank $|z| = 1.81$, $p \leq 0.070$; *Baseline* only: $|z| = 1.78$, $p \leq 0.075$; *Anticipation* only $|z| = 1.69$, $p \leq 0.091$).[33] For any investment, a trustee who makes a relative return transfer of Y=2 is not treated different than a passive player (Wilcoxon signed-rank $|z| = 1.47$, $p \geq 0.142$; *Baseline* only: $|z| = 1.122$, $p \geq 0.224$; *Anticipation* only $|z| = 1.41$, $p \geq 0.160$). For each investment, a relative return transfer of Y>2 is rewarded (Wilcoxon signed-rank $|z| = 2.31$, $p \leq 0.021$; *Base-*

31 Please note that each outsider and each trustee receives an endowment of 100 ECU. Only after receiving an investment from an investor can a trustee become active and thereby increase the investor's and his own payoffs. In case the investor returns the entire transfer, he ends up with the same payoff as an outsider or a passive trustee (with 100 ECU).

32 Note that not giving something can be interpreted as punishing.

33 This result is not supported in the *Anticipation* treatment, if a trustee makes a relative return transfer of Y=1 and has previously received an investment of X≥20.

line only: |z| = 1.65, p ≤ 0.099; *Anticipation* only |z| = 1.73, p ≤ 0.084).[34] Given these observations on can state:

> **Result 1a′:** Helpers' transfers to trustees, who implemented an equal split in the trust game, do not significantly differ from the transfer to passive players. Relative return transfers that leave the investor with more money than the trustee lead to higher helpers' transfers, while lower return transfers lead to lower helpers' transfers.

5.2. Return Transfers

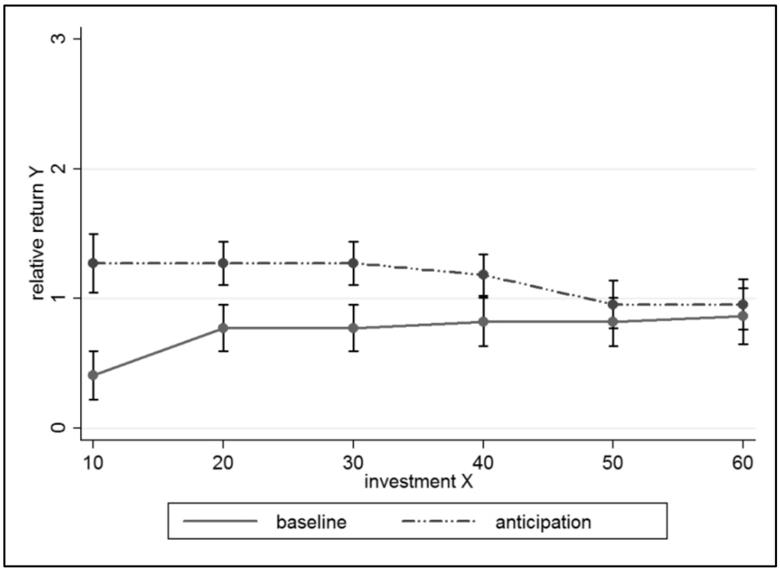
This section analyzes whether trustees correctly anticipate helpers' transfer decisions and therefore reciprocate investors' transfers more in the *Anticipation* treatment than in the *Baseline*. In addition, a positive correlation between investments and return transfers is examined.

Average return transfers are significantly higher in the *Anticipation* treatment than in the *Baseline* (Mann-Whitney rank-sum |z| = 2.14, p = 0.033). In addition, Figure 2 demonstrates a difference in the pattern of transfers between the treatments. In the *Baseline*, a typical outcome for the trust game can be observed (see Falk, Meier, and Zehnder 2013): the higher the investment, the higher the relative return transfer (Spearman's Rho: r_S = 0.15, p = 0.088). In the *Anticipation* treatment, the reverse occurs: lower investments are reciprocated more (Spearman's Rho: r_S = -0.15, p = 0.085).[35] The relative return transfers differ significantly between the treatments for investments lower than 50 (Mann-Whitney rank-sum |z| = 1.68, p ≤ 0.093), while they are not statistically different for the highest two possible investments (Mann-Whitney rank-sum |z| = 0.61, p ≥ 0.545). This indicates that the motives behind the return transfers change in the *Anticipation* treatment compared to the *Baseline*, i.e., trustees "invest" in a reputation of being trustworthy when it is cheap (when a high relative return transfer results in a comparatively low absolute return transfer) and do not reciprocate high investments more than in the *Baseline*. This finding is especially interesting given that studies involving reward by second parties (direct reciprocity) in one-shot, public-good games have not found an increase in socially desirable behavior (e.g., Walker and Halloran, 2004).

34 The only exception is in the *Anticipation* treatment with an investment of X=10.

35 Comparing for each investment the relative return transfers (return transfer for X=10 vs. for X=20; for X=20 vs. for X=30, etc.) overall (and in the *Baseline*), only the first comparison is significantly different (|z| = 1.73, p = 0.084; *Baseline*: (|z| = 2.18, p = 0.029), while the rest is not statistically significantly different (|z| = 1.00, p ≥ 0.317; *Baseline*: (|z| = 1.41, p ≥ 0.157). In the *Anticipation* treatment, only an investment of X=50 is reciprocated significantly more than an investment of X=40 (|z| = 1.65, p = 0.099), while no other comparison is statistically significantly different (|z| = 1.41, p ≥ 0.157).

Figure 2
Relative Return Transfers by Treatment



On the x-axis, the investment the trustee has received is depicted; on the y-axis, mean relative return transfers are displayed. Standard errors are indicated.

A great number of studies have demonstrated that male subjects tend to act in a more strategic way than female subjects (e.g., Eckel and Grossman, 2008 for an overview). Thus, further support for trustees' strategic concerns is given by the fact that the treatment difference is driven by male subjects (Mann-Whitney rank-sum: *Males*: |z| = 2.43, p = 0.015; *Females*: |z| = 0.17, p = 0.867). Here, a random effects ordered Probit regression is used to further investigate the influence factors determining the return transfers.[36]

36 The random effects control for the fact that each trustee takes six decisions (one for each investment) and an ordered Probit model suits these data the best, since trustees can only choose between four different relative return transfer per investment (Y=0, Y=1, Y=2 or Y=3).

Table 2
Explaining trustees' transfers – comparison *Baseline* and *Anticipation* treatment

Random effects ordered Probit regressions ("trustees" as group) Dependent variable: Relative return transfers in the trust game				
	Model 1	Model 2	Model 3	Model 4
Anticipation	2.45***	2.45***	4.16***	4.20***
	(.27)	(.27)	(.50)	(.52)
Investment		-.00	.02***	.02***
		(.00)	(.01)	(.01)
Investment**Anticipation*			-.04***	-.05***
			(.01)	(.01)
Gender**Anticipation*				-1.11**
				(.44)
Cut 1 Constant	.29	.25	1.13***	1.06***
	(.18)	(.25)	(.34)	(.33)
Cut 2 Constant	2.29***	2.25***	3.28***	3.31***
	(.21)	(.27)	(.38)	(.39)
Cut 3 Constant	4.54***	4.50***	5.65***	5.51***
	(.39)	(.41)	(.52)	(.50)
Rho Constant	.86***	.86***	.87***	.86***
	(.02)	(.02)	(.02)	(.02)
N	264	264	264	264
N of groups	44	44	44	44
P model	0.033	0.101	<.001	<.001
LR Chi2	4.52	4.58	23.73	22.74

Random effects ordered Probit regressions. Standard errors are presented in parentheses. Observations are clustered on individual level. The *Anticipation* dummy equals 1 for all observations of the *Anticipation* treatment, *investment* controls for the investment (X) the trustee has received, the *gender* dummy equals 1 for all female observations. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

The most important insight from the models in Table 2 is that controlling for different additional influence variables, the data confirm Hypothesis 2b, i.e., the coefficient of the treatment dummy (*Anticipation*) remains highly significant. At first glance, the investment alone (Model 2) does not appear to have a correlation with the return transfer. Yet, one can see a positive significant correlation of the investment and the return transfer when including the interaction of the investment and the treatment dummy (Models 3). The interaction effect is negative and statistically significant. Including an interaction term between subjects' gender and the treatment (Model 4) one can see that women respond significantly less to the treatment than men. Thus, the results generally support Hypothesis 2a, but the effect is reversed in the *Anticipation* treatment.

Summing up, the results of the random effects ordered Probit regression support the findings from the non-parametric data analysis. Overall, the results support Hypothesis 2b, while Hypothesis 2a can only be supported in the *Baseline*. This leads to:

Result 2a: In the *Baseline*, relative return transfers increase as the investments increase. As opposed to this, relative return transfers are high for low investments and decrease with high investments in the *Anticipation* treatment.

and to:

Result 2b: The return transfers are, on average, higher in the *Anticipation* treatment than in the *Baseline*. The result is mainly driven by male subjects.

5.3. Investments

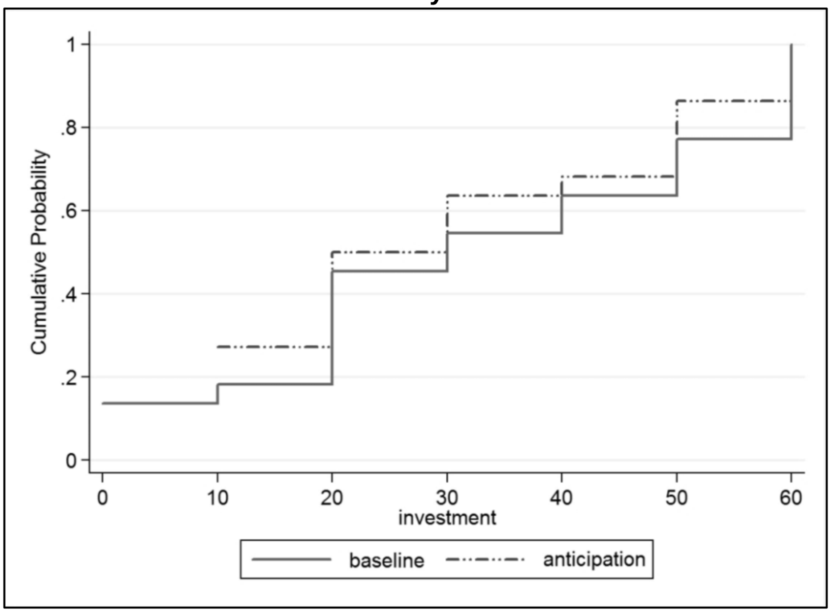
This section tests whether investors correctly anticipate the different return transfers between the treatments and therefore invest more in the *Anticipation* treatment compared to the *Baseline*.

Statistically, the investments do not differ between the treatments (Mann-Whitney rank-sum: $|z| = 0.47$, $p = 0.642$), but the cumulative distribution function of the investments in Figure 3 shows that the distribution of the investments differs. Specifically, only in the *Baseline* but not in the *Anticipation* treatment zero-investments are made.[37]

Nonetheless, the data do not support Hypothesis 3 and lead to:

Result 3: On average, investments do not statistically differ significantly between treatments.

Figure 2
Investments by Treatment



The graph shows an empirical cumulative distribution function of the investments for both treatments. On the x-axis, the investment is depicted; on the y-axis, the estimated probability for each investment is displayed.

37 Using Fisher's exact test for a positive investment does not lead to a statistically significant treatment difference (1-sided Fisher's exact = 0.116). Testing the data parametrically with an ordered Probit regression model does not lead to any significant results either. No interaction between subjects' gender and the treatment can be found.

6. Discussion

As mentioned above, helpers seem to care more about socially desirable outcomes than about the motives behind trustees' transfers. Those preferences raise the question of whether the monetary efficiency is, in fact, higher in the *Anticipation* treatment. As shown above, helpers make, on average, the same transfers in the *Anticipation* treatment as in the *Baseline* given a particular return transfer and trustees anticipate that. I.e., trustees make higher relative return transfers in the *Anticipation* treatment especially if such an "investment" in a good reputation is cheap in absolute terms. It turns out that the money-maximizing strategy for trustees (given the strategy of the helpers) in the *Anticipation* treatment is fully to return (Y=3) each investment (X).[38] Thus, trustees do not anticipate helpers' positive strong indirect reciprocity enough to maximize their profits.

Given the decisions of trustees investors' money-maximizing strategy is to invest X=0 in *Baseline* and X=30 in the *Anticipation* treatment. However, the quantitative difference in payoffs is very small (100 compared to 108.18). If investors are at least slightly risk-averse, they make the correct decision by not investing differently in the *Anticipation* treatment compared to the *Baseline*. An alternative explanation for the underinvestment could be that the level of reasoning is too high. I.e., investors have to anticipate not only what trustees will do, but also what they think what trustees think the helpers will do in the later game.[39]

7. Conclusion

From a welfare point of view, rewards are better for the society than punishment since they do not lead to an efficiency loss. Rand et al. (2009, p. 1272) show in a repeated public goods game that "*reward is as effective as punishment for maintaining public cooperation and leads to higher total earnings. Moreover, when both options are available, reward leads to increased contributions and payoff, whereas punishment has no effect on contributions and leads to lower payoff*". In a world in which anonymous interactions become more and more frequent (e.g., via the internet), it is important to pay more attention to economic consequences of reward systems by impartial strangers. Surprisingly, until now only a small number of papers analyzes rewards rather than punishments and even less literature considers rewards given by non-strategically motivated third parties. Most notably, there exists no literature on the question: How can the anticipation of a reward from an impartial stranger enhance trustworthiness? The current paper attempts to close this gap.

This paper shows that positive strong indirect reciprocity exists. Further, helpers' transfers are surprisingly more positively correlated with relative return transfers in the *Anticipation* treatment than in the *Baseline*. Most importantly, trustees anticipate this behavior correctly and

38 Recall that in the *Baseline*, in expectations the money-maximizing strategy for trustees is to send zero transfers (since in the *Baseline* trustees do not know that the helping game will follow). Similarly, the money-maximizing strategy in both treatments for helpers is to send zero transfer.

39 See the level-k literature (e.g., Costa-Gomes, Crawford, and, Broseta, 2001; Nagel, 1995; Stahl and Wilson, 1995).

become more trustworthy in the *Anticipation* treatment than in the *Baseline*. In particular, they make higher relative return transfers in response to lower investments, signaling higher trustworthiness when such signals are cheap in absolute terms. They would have earned even more than their counterparts in the *Baseline* if they had fully anticipated the complete strategy of helpers.

The experiment still leaves some open questions. There is high zero-censoring in the helping game. This could be interpreted as helpers' willingness to punish greedy trustees. An experiment where real punishment is used instead of denying help may be used to understand the helpers' actions further. Dohmen, Huffman, Sunde (2009) show that positive and negative reciprocity are different concepts. This finding might be true for strong negative and positive indirect reciprocity as well. Without encountering ethical concerns, one could implement only an announced version of the punishment game.

This paper helps to explore the motivations behind and the consequences of third-party reward of trustworthiness. The findings from this paper are applicable to many different settings. The results show that, in all situations in which people display trustworthy behavior for intrinsic reasons, they might in fact do a better job if they could expect strangers to reward them for their actions. That is even true if the strangers do not have any strategic incentives to do so and if their actions are costly. From a policy perspective, it turns out that it can be socially beneficial to promote greater publicity on socially desirable acts.

References

Abeler, J., Falk, A., Goette, L., Huffman, D., 2011. "Reference points and effort provision." *American Economic Review* 101(2), 470–492.

Almenberg, J., Dreber, A., Apicella, C. L., Rand, D. G., 2011. "Third Party Reward and Punishment: Group Size, Efficiency and Public Goods." In N. M. Palmetti, and J. P. Russo (eds.), *Psychology and Punishment*, New York: Nova Science Publishers, 73-92.

Andreoni, J., 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *The Economic Journal* 100(401), 464–477.

Andreoni, J., Miller, J. H., 1993 "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence." *The Economic Journal* 103(418), 570-585.

Arrow, K. J., 1972. "Gifts and Exchanges." *Philosophy and Public Affairs* 1(4), 343-362.

Balafoutas, L., Nikiforakis, N., Grechenig, K., 2014. "Third-party punishment and counter-punishment in one-shot interactions." *Economics Letters* 122(2), 308–310.

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

Bernhard, H., Fischbacher, U., Fehr, E., 2006. "Parochial altruism in humans." *Nature* 442, 912-915.

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

Bardsley, N., 2008. "Dictator Game Giving: Altruism or Artefact?" *Experimental Economics* 11(2), 122-133.

Boyd, R., Richerson, P. J., 1989. "The evolution of indirect reciprocity." *Social Networks* 11, 213–236.

Carpenter, J. P., Matthews, H. P., 2004. "Social reciprocity." *IZA Discussion Paper* No. 1347.

Carpenter, J. P., Matthews, H. P., Ong'ong'a, O., 2004. "Why Punish? Social reciprocity and the enforcement of prosocial norms." *Journal of Evolutionary Economics* 14, 407–429.

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

Costa-Gomes, M., Crawford, V. P., Broseta, B., 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica* 69(5), 1193-1235.

Costa-Gomes, M., Huck, S., Weizsäcker, G., 2014. “Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect.” *Games and Economic Behavior* 88. 298–309.

Cox, J. C., Friedman, D., Sadiraj, V., 2008. “Revealed Altruism.” *Econometrica*, 76(1), 31–69.

Deci E.L., R.M. Ryan, 2000. “Intrinsic and Extrinsic Motivations: Classic Defnitions and New Directions.” *Contemporary Educational Psychology* 25, 54-67.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2009. “Homo Reciprocans: Survey Evidence on Behavioural Outcomes.” *The Economic Journal* 119(536), 592–612.

Dufwenberg, M., Kirchsteiger, G., 2004. “A theory of sequential reciprocity.” *Games and Economic Behavior* 47, 268–298.

Eckel, C.C., Grossman, P. J., 2008. “Differences in the Economic Decisions of Men and Women: Experimental Evidence.” In C.R. Plott and V. L. Smith, (eds.), *Handbook of Experimental Economics Results* 1, Elsevier B.V, 509-519.

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

Falk, A., Fischbacher, U., 2006. “A Theory of Reciprocity.” *Games and Economic Behavior* 54(2), 293-315.

Falk, A., Fehr, E., Fischbacher, U., 2008. “Testing theories of fairness – Intentions matter.” *Games and Economic Behavior* 62(1), 287–303.

Falk, A., Meier, S., Zehnder, C., 2013. “Do Lab Experiments Misrepresent Social Preferences? The Case of Self-Selected Student Samples.” *Journal of the European Economic Association* 11(4), 839-852.

Fehr, E., Fischbacher, U., 2004. “Third-party punishment and social norms.” *Evolution and Human Behavior* 25, 63-87.

Fehr, E., Gächter, S., 2000. “Cooperation and punishment in public goods experiments.” *American Economic Review* 90, 980-994.

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

Fischbacher, U., Gächter, S., Quercia, S., 2012. “The Behavioural Validity of the Strategy Method in Public Good Experiments.” *Journal of Economic Psychology* 33(4), 897-913.

Forsythe, R., Horowitz, J. L., Savin, N. E., Sefton, M., 1994. “Fairness in Simple Bargaining Games.” *Games and Economic Behavior* 6(3), 347–369.

Gintis, H. (2000). “Strong Reciprocity and Human Sociality”. *Journal of Theoretical Biology* 206(2), 169–179.

Greiner, B., 2015. “Subject pool recruitment procedures: Organizing experiments with ORSEE.” *Journal of the Economic Science Association* 1(1), 114-125.

Greiner, B., Levati, M. V., 2005. “Indirect reciprocity in cyclical networks. An experimental study.” *Journal of Economic Psychology* 26, 711–731.

Holt, C. A., Laury, S. K., 2002. “Risk Aversion and Incentive Effects.” *American Economic Review* 92(5), 1644-55.

Levine, D. K., 1998. “Modeling Altruism and Spitefulness in Experiments.” *Review of Economic Dynamics* 1, 593-622.

Nagel, R., 1995. “Unraveling in Guessing Games: An Experimental Study.” *American Economic Review* 85(5), 1313-1326.

Nikiforakis, N., Mitchell, H., 2014. “Mixing the carrots with the sticks: Third party punishment and reward.” *Experimental Economics* 17(1), 1-23.

Nowak, M.A., Sigmund, K., 1998. “Evolution of indirect reciprocity by image scoring.” *Nature* 393, 573–577.

Rabin, M., 1993. “Incorporating fairness into game theory and economics.” *American Economic Review* 83, 1281–1302.

Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., Nowak, M. A., 2009. “Positive Interactions Promote Public Cooperation.” *Science* 325, 1272-1275.

Rockenbach, B., Milinski, M., 2006. “The efficient interaction of indirect reciprocity and costly punishment.” *Nature* 444, 718–723.

Seinen, I., Schram, A., 2006. “Social status and group norms: Indirect reciprocity in a repeated helping experiment.” *European Economic Review* 50, 581–602.

Selten, R., 1967. “Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments.” In E. Sauermann (ed.), *Beiträge zur experimentellen Wirtschaftsforschung*, Tübingen, Mohr 136-168.

Servátka, M., 2009. “Separating reputation, social influence, and identification effects in a dictator game.” *European Economic Review* 53, 197-209.

Stahl, D. O., Wilson, P. W., 1995. “On Players’ Models of Other Players: Theory and Experimental Evidence.” *Games and Economic Behavior* 10(1), 218-254.

Stanca, L., 2009. “Measuring Indirect Reciprocity: Whose Back Do We Scratch?” *Journal of Economic Psychology* 30(2), 190-202.

Stanca, L., Bruni, L., Corazzini, L., 2009. “Testing Theories of Reciprocity: Do Motivations Matter?” *Journal of Economic Behavior & Organization* 71(2), 233-245.

Walker, J. M., Halloran, M., 2004. “Rewards and Sanctions and the Provision of Public Goods in One-Shot Settings.” *Experimental Economics* 7(3), 235-247

Appendix A. Tables and Figures

Table 1A
Descriptive statistics of choices in all games

		Investors' Choices (in %)		*Trustees' Choices (in %)*		*Mean Helpers' Choices (standard errors in parentheses)*	
		Baseline	*Anticipation*	*Baseline*	*Anticipation*	*Baseline*	*Anticipation*
Outsider						11.64 (22.83)	9.14 (10.57)
Investment (X)	*Relative Return Transfer (Y)*						
0		13.64	0.00			12.59 (22.72)	10.95 (14.82)
10	0	4.55	27.27	77.27	27.27	0.73 (2.33)	1.82 (5.88)
	1			9.09	31.82	3.73 (6.36)	3.73 (4.50)
	2			9.09	27.27	12.68 (21.72)	9.64 (8.06)
	3			4.55	13.64	15.00 (21.93)	14.45 (15.52)
20	0	27.27	22.73	40.91	13.64	0.95 (2.94)	2.27 (10.66)
	1			50.00	50.00	3.00 (6.28)	6.27 (12.38)
	2			9.09	31.82	12.36 (22.35)	14.55 (15.58)
	3			0.00	4.55	14.14 (22.31)	19.23 (21.59)
30	0	9.09	13.64	40.91	13.64	0.50 (2.13)	0.55 (2.15)
	1			45.45	50.00	3.00 (5.47)	7.27 (12.33)
	2			9.09	31.82	12.82 (22.40)	14.09 (17.57)
	3			4.55	4.55	15.82 (23.43)	22.27 (27.31)
40	0	9.09	4.55	40.91	13.64	0.95 (2.94)	0.45 (2.13)
	1			40.91	59.09	4.23 (8.48)	7.64 (12.15)
	2			13.64	22.73	13.95 (24.22)	18.41(25.70)
	3			4.55	4.55	15.14 (22.96)	26.55 (32.42)
50	0	13.64	18.18	40.91	31.82	1.86 (6.64)	0.68 (2.34)
	1			40.91	45.45	4.23 (8.48)	8.55 (15.08)
	2			13.64	18.18	12.73 (22.42)	20.09 (27.05)
	3			4.55	4.55	15.00 (22.41)	25.77 (30.95)
60	0	22.73	13.64	36.36	50.00	2.32 (8.68)	0.91 (4.26)
	1			36.36	18.18	4.68 (9.06)	8.09 (16.25)
	2			22.73	27.27	13.59 (22.57)	19.09 (26.03)
	3			4.55	4.55	16.09 (23.22)	23.82 (27.08)

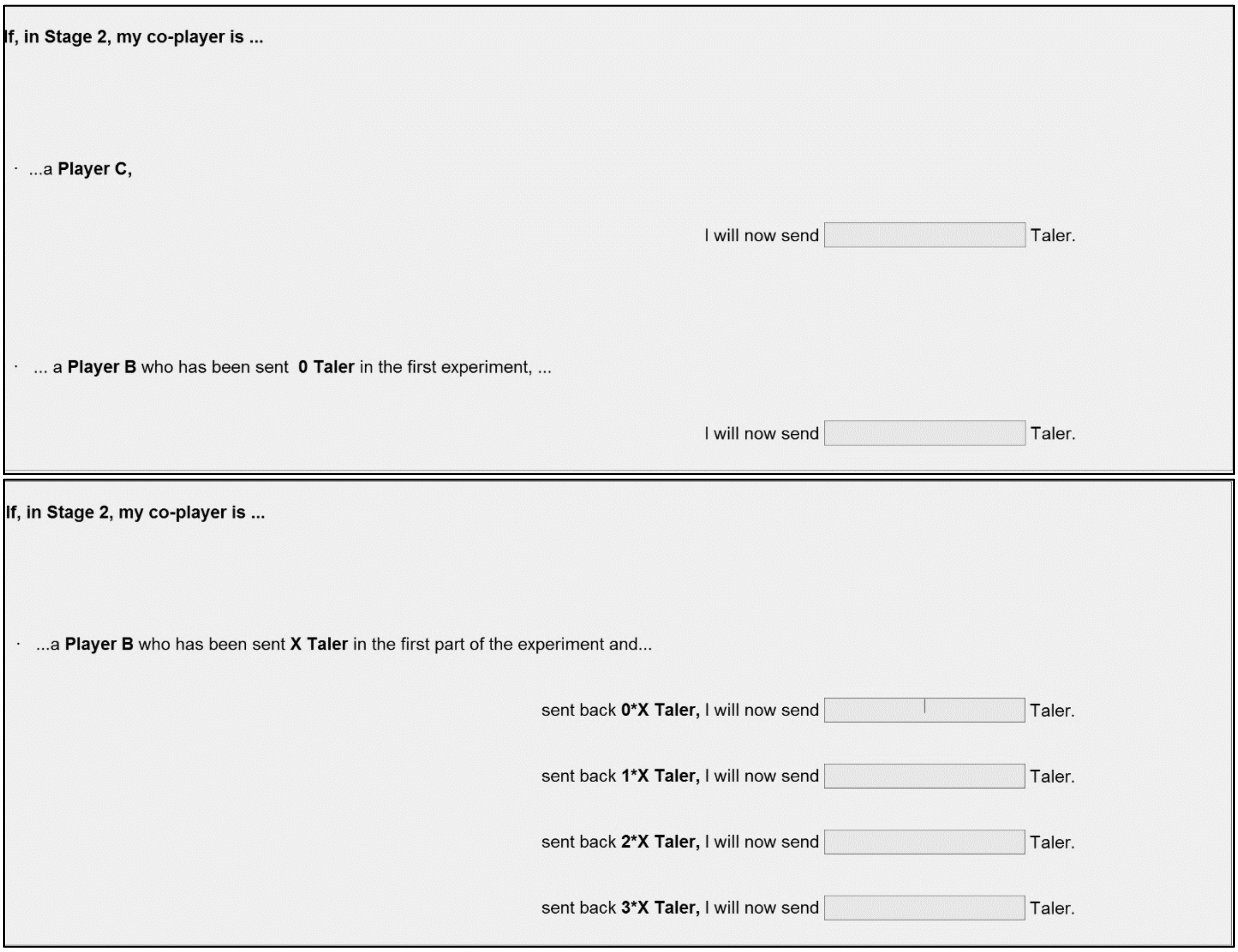
N = 22 in each column
The distributions of investors' choices and trustees' choices conditional on investments are depicted in columns 3-6. Columns 7-8 present mean helpers' transfers conditional on the role of the recipient (outsider

Figure 1A
Screenshot – Trustees' Strategy

In case player A has sent me...	
... **0 Taler,** my income has thereby increased by **0 Taler.**	
... **10 Taler** and my income has thereby increased by **30 Taler,** I will now send him ...	nothing (0) the transfer (10) double the transfer (20) triple the transfer (30)
... **20 Taler** and my income has thereby increased by **60 Taler,** I will now send him ...	nothing (0) the transfer (20) double the transfer (40) triple the transfer (60)
... **30 Taler** and my income has thereby increased by **90 Taler,** I will now send him ...	nothing (0) the transfer (30) double the transfer (60) triple the transfer (90)
... **40 Taler** and my income has thereby increased by **120 Taler,** I will now send him ...	nothing (0) the transfer (40) double the transfer (80) triple the transfer (120)
... **50 Taler** and my income has thereby increased by **150 Taler,** I will now send him ...	nothing (0) the transfer (50) double the transfer (100) triple the transfer (150)
... **60 Taler** and my income has thereby increased by **180 Taler,** I will now send him ...	nothing (0) the transfer (60) double the transfer (120) triple the transfer (180)

The screen for the elicitation of trustees' choices via the strategy method is depicted. In the first column, the trustee can see how high the investment could have been. In the second column, the participant sees a radio button, on which he can click – for each possible investment X>0, he can choose how much he wants to send back to his investor, i.e., he can choose for each investment X>0 his return transfer X*Y, where Y can be 0 ("nothing"), 1 ("the transfer"), 2 ("double the transfer"), or 3 ("triple the transfer").

Figure 2A
Screenshot – Helpers' Strategy



In the upper box, the first screen of the elicitation of helpers' choices via the strategy method is depicted. Here, the helper can choose how much she wants to send to the recipient in case the recipient is an outsider (called "Player C" in the instructions) or in case the recipient is a passive trustee (called "Player B" in the instructions who has not received an investment). The lower box shows the remaining screens of the helpers' strategy method. Here, the helper can choose how much she wants to transfer to the recipient in case he is a trustee and has received an investment of X (X=10 on the 2nd screen, X=20 on the 3rd screen, etc.) and in case he has then sent back Y*X (Y can be 0, 1, 2, or 3). On the actual screens of the players, instead of "X", "0*X", or "1*X" etc., the absolute numbers of the respective transfers are written. The helper can insert in each line a number between 0 and 100. Only the transfer decision for the relevant situation will become payoff-relevant.

Appendix B. Instructions

Here, the English translation of the instructions is presented. The original German version can be passed on upon request. The only difference between the *Anticipation* treatment and the *Baseline* is that, in the *Anticipation* treatment, first only the general instructions and the instructions for the trust game are presented and only after the trust game is player, the instructions for the helping game are handed out. On the contrary, in the *Baseline* the both instructions are handed out immediately at the beginning of the experiment. The instructions and the control questions are presented here in the order in which they are read out aloud by the experimenter (and filled out by the participants).

General Instructions for Participants

You are about to take part in economic experiments. Depending on the decisions that you and others make, you can earn a substantial amount of money. It is therefore very important that you read these instructions carefully.

The written statements you have received from us serve your own private information only. **During the experiments, any communication whatsoever is forbidden.** If you have any questions, please ask only us. Please raise your hand and we will come to you. Disobeying these rules will lead to exclusion from the experiments and from all payments.

During the experiments, we speak not of Euro, but instead of Taler. Your entire income is hence initially calculated in Taler. The total number of Taler you earn during the experiments is converted into Euro at the end and paid to you in cash, at the rate of

1 Taler = 2 Eurocent.

In addition, each participant is paid a lump sum of 4 Euro for showing up today.

You will take part in several experiments today. The instructions to each experiment will be handed out to you one by one, just before the respective experiment is about to begin. On the following pages, we will describe the exact procedure of the first experiment.

In this experiment, there are three different roles: Player A, Player B, and Player C. At the beginning of the experiment, you are randomly assigned one of these three roles.

In the experiment, you are required to make your decision once only; i.e., the experiment is conducted only once. You will thus make no repeated decisions today.

Information on the **First Experiment**

In this experiment, one Player A and one Player B are randomly paired. Each player initially receives 100 Taler, which we shall refer to in the following as the initial endowment.

This experiment consists of 2 stages.

Stage 1: In the first stage, **Player A** decides which transfer of X Taler from the initial endowment he or she wishes to send to B. X may be one of the following values: 0, 10, 20, 30, 40, 50, or 60. The Taler sent by Player A are **tripled** and transferred to the Player B who has been assigned to him/her.

Stage 2: If Player A has sent more than 0 Taler in Stage A, **Player B** now has to decide, in Stage 2, how many Taler Y*X he or she wishes to transfer back to A. Player B may choose from the following options: no Taler (Y = 0), the same amount of Taler as in the transfer (Y = 1), double the transferred amount of Taler (Y = 2), or three times the transferred amount of Taler (Y = 3).

The income at the end of the first part of the experiment is therefore:

Player A:	
If he/she has sent 0:	100
If he/she has sent X:	100 - X + (X * Y)
Player B:	
If he/she receives 0:	100
If he/she receives X:	100 + 3 * X - (X * Y)

Here, X may be equal to 0, to 10, to 20, to 30, to 40, to 50, or to 60.
Y may be equal to 0, to 1, to 2, or to 3.

Examples:

1.) If, for example, Player A sends a sum of X = 10 Taler and B decides to send back three times the amount of the transfer (Y = 3), then both players have the following income:

A: 100 - 10 + (10 * 3) = 120, and B: 100 + 3 * 10 - 3 * 10 = 100.

2.) If, for example, Player A sends a sum of X = 60 Taler and B decides to send back the amount of the transfer (Y = 1), then both players have the following income:

A: 100 - 60 + (60 * 1) = 100, and B: 100 + 3 * 60 - 1 * 60 = 220.

3.) If, for example, Player A sends a sum of X = 40 Taler and B decides to send back twice the amount of the transfer (Y = 2), then both players have the following income:

A: 100 - 40 + (40 * 2) = 140, and B: 100 + 3 * 40 - 2 * 40 = 140.

However, only at the end of Stage 2 are you told how high your income is from the experiment.

Player C does not have to make any decision in this experiment. Player C's payoff in this experiment is 100 Taler.

The experiment ends here. You are then told how high your income is from the experiment. Further experiments follow – **however, it is impossible for you to be assigned once again to a group with the same players**. **Further, you cannot lose your payoffs from the experiments**. Following the final experiment, you will be given a questionnaire. Once you have filled in the questionnaire, you will receive your payoff from us in cash. In order to receive your payoff, please bring all documents you have received from us with you.

Information on the **Second Experiment**

Participants in this experiment have the same roles as in the first experiment. Hence, this means that a participant who had role A in the first experiment will still have role A in this experiment; a participant who had role B in the first experiment will still have role B in this experiment; and a participant who had role C in the first experiment will still have role C in this experiment. Random assignation is conducted anonymously, and we ensure that **you are not assigned once again to a group with the same players as in the first experiment**.

This second experiment also consists of two stages.

Stage 1: In the first stage, **Player A** receives an endowment of 100 Taler. He or she now has to decide how many of these 100 Taler to send his or her co-player (full numbers between 0 and 100). Every Taler sent is **tripled** and credited to the other player's account.

Whether the co-player is a Player B or a Player C is decided randomly only in Stage 2. However, Player A must make a binding decision in Stage 1 on how many Taler to send if the co-player is a Player B, and how many Taler to send if the co-player is a Player C. In addition, if it is a Player B, Player A may also decide how many Taler to send depending on Player B's behavior in the first experiment. In Stage 1, Player A therefore has to fill in the following seven decision tables for all possible transfers *X* and *Y* * *X*:

This is what the decision tables look like:

1st Screen

If, in Stage 2, my co-player is …

- … a **Player C**, …

  … I will now send _____Taler.

- … a **Player B** who has been sent **0 Taler** in the first experiment, …

  … I will now send _____Taler.

2nd Screen

If, in Stage 2, my co-player is …

- … a **Player B** who has been sent **10 Taler** in the first experiment and…

… sent back **0** **Taler**, I will now send _____Taler.
… sent back **10** **Taler**, I will now send _____Taler.
… sent back **20** **Taler**, I will now send _____Taler.
… sent back **30** **Taler**, I will now send _____Taler.

3rd Screen

If, in Stage 2, my co-player is …

- …a **Player B** who has been sent **20 Taler** in the first experiment and …

… sent back **0** **Taler**, I will now send _____Taler.
… sent back **20** **Taler**, I will now send _____Taler.
… sent back **40** **Taler**, I will now send _____Taler.
… sent back **60** **Taler**, I will now send _____Taler.

4th Screen

If, in Stage 2, my co-player is …

- …a **Player B** who has been sent **30 Taler** in the first experiment and …

… sent back **0** **Taler**, I will now send _____Taler.
… sent back **30** **Taler**, I will now send _____Taler.
… sent back **60** **Taler**, I will now send _____Taler.
… sent back **90** **Taler**, I will now send _____Taler.

5th Screen

If, in Stage 2, my co-player is …

- … a **Player B** who has been sent **40 Taler** in the first experiment and …

  … sent back **0 Taler**, I will now send _____Taler.
  … sent back **40 Taler**, I will now send _____Taler.
  … sent back **80 Taler**, I will now send _____Taler.
  … sent back **120 Taler**, I will now send _____Taler.

6th Screen

If, in Stage 2, my co-player is …

- … a **Player B** who has been sent **50 Taler** in the first experiment and …

  … sent back **0 Taler**, I will now send _____Taler.
  … sent back **50 Taler**, I will now send _____Taler.
  … sent back **100 Taler**, I will now send _____Taler.
  … sent back **150 Taler**, I will now send _____Taler.

7th Screen

If, in Stage 2, my co-player is …

- … a **Player B** who has been sent **60 Taler** in the first experiment and …

  … sent back **0 Taler**, I will now send _____Taler.
  … sent back **60 Taler**, I will now send _____Taler.
  … sent back **120 Taler**, I will now send _____Taler.
  … sent back **180 Taler**, I will now send _____Taler.

You may enter any number between 0 and 100 in each line. It goes without saying that the lines do **not** have to add up to 100 either, for **only that line is decision-relevant which actually corresponds to the situation drawn by lot in Stage 2!** The decisions in the other lines (not drawn) do not influence your payoff. Do please note, however, that while filling in the table you do not yet know if your co-player is a Player C or B (and you do not know, in case a Player B is assigned to you, how the he has behaved in the first part of the experiment). **In each line, you therefore have to consider your decision carefully, for every one can become relevant for you.**

Stage 2: In the second stage, it is decided by draw whether Player A's co-player is a Player B or C. The player who is drawn then receives the amount of Taler according to the corresponding decision table from Stage 1. He or she therefore does not have to make any decision in this second experiment.

The other player – who is not drawn – is given the possibility to increase his or her own income individually, by way of a small task. (Precise instructions for this task will appear later on this player's screen.)

The income from the second experiment is therefore:

Player A:
100 - (what A sent the co-player)

Co-player (B or C, depending on the draw in Stage 2):
3 * (what A sent the co-player)

Player not drawn (B or C, depending on the draw in Stage 2):
Income from the individual small task

Examples:

1.) If Player A should hence decide to enter the following numbers in the second table: top line 1, second line 13, third line 17, bottom line 0; and if the co-player assigned to Player A is a Player B who was sent 10 Taler by his co-player in the first experiment, and who in turn decided to send nothing back, then the payoffs from this second experiment are as follows for Player A, who made the decision here:

100 - 1 = 9. The payoff for the co-player assigned to Player A is: 1 * 3 = 3.

2.) If Player A should decide to enter the following numbers in the first table: top line 99, second line 14; and if the co-player assigned to Player A is a Player C, then the payoffs from this second experiment are as follows for Player A, who has made the decision here:

100 - 99 = 1. The payoff for the co-player assigned to Player A is: 99 * 3 = 297.

The experiment ends here. You will then hear about your payoff from the first and second experiment. Further experiments will follow – **however, it is impossible for you to be assigned once again to a group with the same players**. **Further, you cannot lose your payoffs from the experiments**. Following the final experiment, you will be given a questionnaire. Once you have filled in the questionnaire, you will receive your payoff from us in cash. In order to receive your payoff, please bring all documents you have received from us with you.

Information on the **Third Experiment**

In this part of the experiment, no other participant is paired with you. The payoffs therefore relate only to you. The decisions of the other participants only have an influence on their own respective payoffs.

In this part of the experiment, you are asked to decide in 10 different situations (lotteries) between option A and B. These situations will be presented to you on consecutive screens. The two lotteries each comprise 2 possible monetary payoffs, one high and one low, which will be paid to you with different probabilities.

The options A and B will be presented to you on the screen, as in the following example:

The computer uses a random draw program, which assigns you payments exactly according to the denoted probabilities.

For the above example, this means:
Option A obtains a payoff of 2 Euro with a probability of 10% and a payoff of 1.60 Euro with a probability of 90%.
Option B obtains a payoff of 3.85 Euro with a probability of 10% and a payoff of 0.10 Euro with a probability of 90%.
Now you have to click on the particular option you decide for.

Please note that, at the end of the experiment, only one of the 10 situations will eventually be paid. Yet, each of the situations can be randomly chosen with equal probability to be the payoff-relevant one.
After this, a draw will determine whether for the payoff-relevant situation the high payoff (2.00 Euro or 3.85 Euro) or the low payoff (1.60 Euro or 0.10 Euro) will be paid.