

Matthey, Astrid; Regner, Tobias

**Working Paper**

## Do reciprocators exploit or resist moral wiggle room? An experimental analysis

Jena Economic Research Papers, No. 2015-027

**Provided in Cooperation with:**

Friedrich Schiller University Jena, Faculty of Economics and Business Administration

*Suggested Citation:* Matthey, Astrid; Regner, Tobias (2015) : Do reciprocators exploit or resist moral wiggle room? An experimental analysis, Jena Economic Research Papers, No. 2015-027, Friedrich Schiller University Jena, Jena

This Version is available at:

<https://hdl.handle.net/10419/144897>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# JENA ECONOMIC RESEARCH PAPERS



# 2015 – 027

## **Do reciprocators exploit or resist moral wiggle room? An experimental analysis**

by

**Tobias Regner  
Astrid Matthey**

[www.jenecon.de](http://www.jenecon.de)

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University Jena, Germany. For editorial correspondence please contact [markus.pasche@uni-jena.de](mailto:markus.pasche@uni-jena.de).

Impressum:

Friedrich Schiller University Jena  
Carl-Zeiss-Str. 3  
D-07743 Jena  
[www.uni-jena.de](http://www.uni-jena.de)

© by the author.

# Do reciprocators exploit or resist moral wiggle room? An experimental analysis

Tobias Regner <sup>♡</sup>\*Astrid Matthey <sup>♣</sup>

<sup>♡</sup> *University of Jena, Jena, Germany*

<sup>♣</sup> *Max Planck Institute of Economics, Jena, Germany*

## Abstract

By now there is substantial experimental evidence that people make use of ‘moral wiggle room’ (Dana et al., 2007), that is, they tend to exploit moral excuses for selfish behavior. However, this evidence is limited to dictator games. In our experiment, a trust game variant, we study whether moral wiggle room also prevails, when reciprocity is a potential motivation for being generous. Trustees’ back transfer choices are elicited for five different transfer levels of the trustor. Moreover, we ask trustees to provide their back transfer schedule for different scenarios that vary the implementation probability of the back transfer. This design allows us to identify subjects who reciprocate and analyze how these reciprocators respond to the provision of moral wiggle room. Our results suggest that moral wiggle room exists as well in the context of reciprocity. Among our subjects, 40% of the reciprocators exploited moral wiggle room.

*JEL classifications:* C72, C91, D03, D80

*Keywords:* social preferences; pro-social behavior; experiments; reciprocity; moral wiggle room; self-image concerns

---

\*Corresponding author: tobias.regner@uni-jena.de

We would like to thank audiences at IMEBE in Madrid, the University of Giessen, the University of Grenoble and at the ESA meeting in Prague for their feedback. We are grateful to Paolo Crosetto for valuable comments. Adrian Liebtrau provided excellent research assistance.

## 1 Introduction

Many people behave pro-socially – if the only other choice is selfish behavior. But what if the situation is less transparent? What if circumstances exist that allow a selfish choice while simultaneously a pro-social image (in front of one self and/or others) can be kept? Dana et al. (2007) find that giving rates are significantly reduced when moral excuses for selfish behavior are available. People seem to make use of ‘moral wiggle room’, a term coined by them, and evidence from a series of studies (e.g., Larson and Capra, 2009; Haisley and Weber, 2010; Hamman et al., 2010; Matthey and Regner, 2011; Lazear et al., 2012; van der Weele, 2013; Grossman, 2014) is in line with this concept. However, this evidence is from dictator games, a setting in which giving has been found to be easily affected by variations of the context or framing.<sup>1</sup> What if ties between giver and taker are stronger? Would moral wiggle room prevail if the relationship is not merely between a dictator and recipient but embedded in a richer environment, say, with social interaction between the two? For instance, having been trusted in the first place may introduce a motivation (reciprocity) strong enough to overcome the trustee’s tendency to exploit moral wiggle room.

Hence, the aim of this paper is to explore whether moral excuses constrain the preference to reciprocate and if so to what extent. For this purpose we conduct a modified trust game. Trustees’ back transfer choices are elicited for five different transfer levels of the trustor. Moreover, we ask trustees to provide their back transfer schedule for different scenarios. While in scenario 1 the back transfer will be implemented for sure, in scenarios 2 to 4 there is a positive probability that the back transfer fails. In such a case the trustee gets to keep the available amount. After trustees have chosen their back transfer schedules for all scenarios, they are informed that they can select the scenario they would like to get implemented.

This design allows us to identify subjects who reciprocate (based on the back transfer schedule in scenario 1) and analyze how these reciprocators respond to the provision of moral wiggle room. Two situational excuses for selfish behavior are present in our design. First, the fact that in scenarios 2 to 4 the transfer could fail may serve as an

---

<sup>1</sup>See, e.g., Hoffman et al. (1996), Cherry et al. (2002), List (2007), Bardsley (2008), Guala and Mittone (2010), Franzen and Pointner (2012).

excuse to return less (or even nothing) in these scenarios. Second, the scenario choice implies the temptation of picking a favorable scenario (for those subjects who returned the same positive amount independently of the scenario) since this means a monetary gain for them (in expectations) while they risk that the trustor they play with does not receive anything. Trustees' choices give us an indication to what extent individuals are willing to use situational excuses for selfish behavior in the context of reciprocity.

Our results suggest that reciprocators make use of moral wiggle room if situational excuses exist. Among our subjects 47 of 117 (40%) reciprocators exploited moral wiggle room, while 70 (60%) resisted. These findings of substantial moral wiggle room among reciprocators are in line with the existing empirical evidence from dictator games. However, they contrast the conclusions of van der Weele et al. (2014), the only other study we are aware of that investigates moral wiggle room in the context of reciprocity.<sup>2</sup> Van der Weele et al. (2014) use a between-subjects design and apply the 'plausible deniability' treatment from Dana et al. (2007) to second-mover behavior in a trust/moonlighting game. Compared to a baseline they find no behavioral differences and conclude that moral wiggle room has no effect on the incidence of reciprocal behavior.

The paper is organized as follows. In section 2 we describe the experiment and present behavioral predictions. Results are reported and discussed in section 3. We conclude in section 4.

## **2 Experiment**

### **2.1 Design**

The experiment consisted of a variant of the trust game (Berg et al., 1995). Both trustor and trustee received an endowment of 10 Euro. As the first step, the trustor could send either 0, 2.50, 5, 7.50 or 10 Euro to the trustee. This transfer was tripled and added

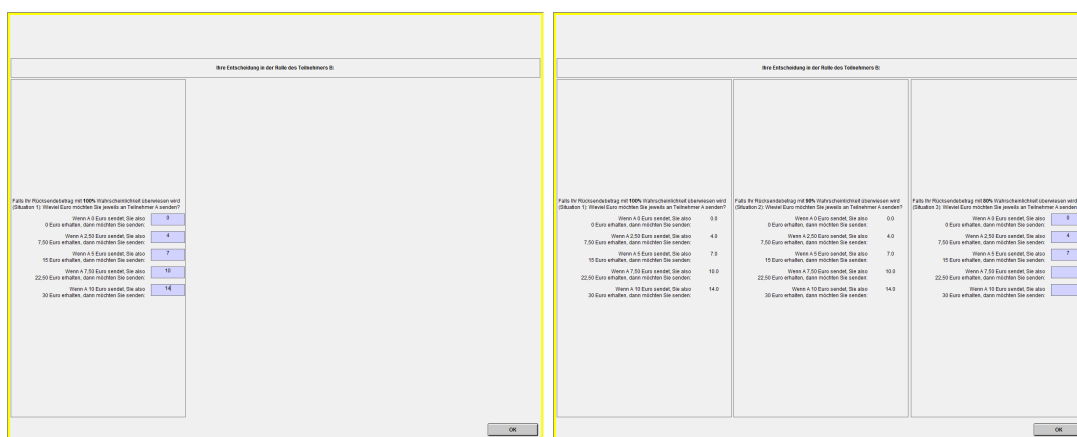
---

<sup>2</sup>Two studies investigate related questions. Conrads and Irlenbusch (2013) analyze strategic ignorance in ultimatum bargaining. They find that unequal proposals in an ultimatum game are rejected less often, if the proposer chose to be ignorant of the payoffs for the responder. Malmendier et al. (2014) analyze the effect of a sorting out option in a double dictator game. They find a high degree of sorting out in a positive reciprocity condition.

to the trustee's account, who could then return any amount available on the account to the trustor. That is, depending on the trustor's transfer trustees could return up to 10, 17.50, 25, 32.50 or 40 Euro. All subjects played in both roles. They knew that it was determined randomly at the end of the experiment whether a subject acted as trustor or trustee. Trustees' decisions were elicited using the strategy method, that is, a trustee decided how much to send back to the trustor for all possible transfers. Hence, all trustees made five back transfer decisions, one of which was to become relevant according to the trustor's actual transfer. When entering their back transfer choices, trustees were informed about the respective amount they would receive at each transfer level. Trustors only learned the outcome, not the choice of the trustee.

Trustees knew that they make the back transfer choices for different scenarios. In scenario 1, the trustee's transfer was carried out with certainty, that is, it reached the trustor for sure and was subtracted from the trustee's account. In scenario 2, the transfer was carried out with 90% probability. With the remaining 10% probability, the trustee would keep the available amount. In this case, the trustor would be left with her endowment minus the amount she sent to the trustee, independently of the size of the trustee's back transfer. In scenario 3, the trustee's transfer was carried out with 80% probability, with 20% the trustee kept the entire amount. In half of our sessions we added a fourth scenario in which the trustee's transfer was or was not carried out with equal probability. Scenario 4 was employed to test whether the availability of an option with a much smaller transfer probability would serve as an excuse to choose a scenario with a transfer probability below 1 (but above 50%) rather than the certain transfer.

Overall, subjects therefore made five back transfer decisions (for each possible amount sent by the trustor) per scenario. After trustees completed all choices for one scenario, they were asked for their back transfers in the next scenario. Choices from previous scenarios were still visible. Figure 1 (a) shows a screenshot of the decision interface for the scenario 1 choice and (b) one for the scenario 3 choice when a subject has already entered back transfers for scenarios 1 and 2. It illustrates the sequential nature of entering the back transfer schedules for the scenarios and the fact that subjects were reminded of their choices in previous scenarios. We chose to provide choices in all previous scenarios in case a subject would like to take the same decision across scenarios.



(a) Scenario 1

(b) Scenario 3

Figure 1: Screenshots of the decision interface (trustees enter their back transfer schedules subsequently for the scenarios with previous choices still visible)

Since not just one decision but an entire back transfer schedule consisting of five choices would have to be remembered, we decided the interface should provide a reminder.

Subjects were instructed that the scenario to be implemented ‘would be decided’ after they made all choices. No specific decision mechanism was mentioned. After subjects had made all decisions, they were shown an overview screen with their transfers for all scenarios and were informed that they could choose themselves which scenario they wanted to apply. Hence, they had the chance to decide whether their transfer would reach the trustor with certainty or not. Finally, we asked subjects a set of additional questions on general dispositions and socio-demographics in a post-experimental questionnaire.

## 2.2 Behavioral Predictions

Assuming pure self-interest the unique subgame perfect Nash equilibrium of our game predicts that the trustee never returns any positive amount. Therefore, the trustor, anticipating this, does not transfer anything. Subjects with reciprocal concerns (Dufwenberg and Kirchsteiger, 2006; or Falk and Fischbacher, 2006) may choose to send/return positive amounts. Based on existing evidence from trust games we expect that a substantial amount of subjects decides to reciprocate. More specifically, we expect that

some trustees return positive amounts when the back transfer is certain, and weakly increase their back transfer with the amount received.

Given subjects reciprocate, we are interested in the way they behave when moral wiggle room exists. Empirical evidence from dictator games suggests that people tend to behave more pro-social, the more salient the relationship between their choice and someone else's outcome is. Essentially, they tend to exploit situational excuses if they are available.

Models of self-image concerns provide a theoretical basis for this behavior, see the literature on cognitive dissonance (Festinger, 1957; Aronson, 1992; Beauvois and Joule, 1996; Konow, 2000), identity (Akerlof and Kranton, 2000; 2005), self-concept maintenance (Mazar et al., 2008), self-signaling (Bénabou and Tirole, 2011). They vary in their approach and terminology, yet their common message is that individuals desire to maintain a comfortable self-image. Deviating from their self-image is costly (in a psychological sense), but the monetary gain of a selfish action may outweigh that cost. For instance, in the self-signaling model of Bénabou and Tirole (2011), situational excuses provide inferential wiggle room and allow individuals to attribute their selfish action to the context, instead of having to connect selfish behavior to one's type. Besides inferential wiggle room (the possibility that inferences from actions about one's type are malleable), Bénabou and Tirole (2011) also allow for another self-signaling process called identity management (beliefs about one's type are malleable through actions). We believe that in the context of our experiment the inferential wiggle room channel of self-signaling is more relevant than Bayesian self-signaling à la identity management.<sup>3</sup>

Generally, decreased salience of a situation would make it easier to engage in self-deception (in cognitive dissonance language) or provides inferential wiggle room (in self-signaling speak). As a consequence, individuals with a desire not to appear *selfish towards themselves* would tend to exploit situational excuses if available. People may also have a desire not to appear *selfish towards others* (see, e.g., Bénabou and Tirole, 2006, or Andreoni and Bernheim, 2009). In that case, decreased salience could make it easier to engage in other-deception and as a result available moral wiggle room would

---

<sup>3</sup>Note that Grossman (2015) models self-image concerns in the identity management way and tests the model experimentally in a binary probabilistic dictator game. His Bayesian self-signaling model predicts that transfers would decrease with the implementation probability. However, results do not offer conclusive supporting evidence.



be exploited by them. As with self-signaling, the behavioral process we focus on is the availability of an excuse that ‘allows’ selfish behavior not the possibility to impress others. Either way (selfish towards themselves or others), reduced transparency would ultimately make it more likely that a selfish action is taken.

Our experimental design offers moral wiggle room to subjects at two stages. First, trustees could exploit the fact that the back transfer is not executed for sure in scenarios 2 to 4.<sup>4</sup> Trustees may use the possible failure of the back transfer as an excuse to *return less* in comparison to their scenario 1 back transfer. They may tell themselves that their transfer may fail anyways and their choice will not matter for the trustor. Hence, the situation in scenarios 2 to 4 allows trustees with a desire not to appear selfish towards themselves to engage in self-deception. Essentially, their self-serving interpretation of the scenario’s risk allows them to be more selfish.<sup>5</sup> Trustees may also have a desire not to appear selfish to others and, hence, may care about the effect of their choice on the trustor. The positive chance of a transfer failure in scenarios 2 to 4 allows them to *return nothing* as the trustor could not distinguish whether getting zero is the consequence of the trustee’s choice or due to the failure of the transfer. Thus, returning nothing in scenarios 2 to 4 is compatible with an image of not appearing selfish to others. The reasoning follows Andreoni and Bernheim (2009). In their 50-50 norm model, the choice of nature (e.g. zero) becomes more attractive with a positive chance of transfer failure and the choice of the equal split shrinks. Of course, in the context of our game the 50-50 norm does not necessarily lead to an equal split but is subjectively interpreted by trustees based on their individual tendency to reciprocate. Moreover, effects in our anonymous setting are expected to be smaller in comparison to the public setting in Andreoni and Bernheim (2009) who analyze audience effects. Overall, the wiggle room concerning the desire not to appear selfish (towards oneself or others) should be more pronounced the lower the implementation probability of the back transfer is.

---

<sup>4</sup>Generally, how much trustees return in scenarios 2 to 4 is not only affected by their social but also by their risk preferences (with respect to their own payoff and the trustor’s). In a recent survey, Trautmann and Vieider (2012) review the literature on social influences on risk attitudes. They find no consistent patterns when risky decisions are taken on behalf of others.

<sup>5</sup>See Haisley and Weber (2010) who find such a self-serving bias due to uncertainty in a related study involving dictator game choices under ambiguity.

**Hypothesis 1** *In comparison to scenario 1 reciprocating trustees tend to transfer back less when the transfer could fail (scenarios 2 to 4).*

Second, moral wiggle room is provided when subjects are informed that they can choose a scenario themselves. Given equal positive back transfers across scenarios, this choice implies a trade-off between the original scenario 1 that implements the back transfer for sure and a scenario that is favorable to the trustee since the transfer may fail. If this moral wiggle room affects the decision of reciprocators, a substantial amount of reciprocating trustees chooses a scenario that involves uncertainty with respect to the implementation of the back transfer.

**Hypothesis 2** *When the choice of a scenario that involves uncertainty results in an expected monetary gain not all reciprocating trustees select scenario 1.*

The choice of a scenario that involves uncertainty may be due to the manipulation's effect on the desire not to appear selfish towards oneself or towards others. Thus, our design cannot distinguish between the two at this stage. As the instructions do not explicitly mention that the chosen scenario is not communicated to the trustor, we cannot rule out that some trustees falsely believed trustors will be informed about the scenario they chose. This would eliminate the situational excuse for trustees motivated by a desire not to appear selfish to others. Only the situational excuse that affects the desire not to appear selfish towards oneself would remain.

### 2.3 Participants and Procedures

128 participants were recruited among students from various disciplines at the local university using the ORSEE software (Greiner, 2004). In each session gender composition was approximately balanced and subjects took part only in one session. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007) and took, on average, 60 minutes. The average earnings in the experiment have been €14.17 (including a €2.50 show-up fee). Sessions took place from November 2011 to January 2012.

Upon arrival at the laboratory subjects were randomly assigned to one of the computer

terminals. Each computer terminal is in a cubicle that does not allow communication or visual interaction among the participants. Participants were given time to privately read the instructions and were allowed to ask for clarifications. In order to check the understanding of the instructions subjects were asked to answer a set of control questions. After all subjects had answered the questions correctly the experiment started. At the end of the experiment subjects were paid in cash according to their performance. Privacy was guaranteed during the payment phase.

### **3 Results**

Our analysis starts with a big picture look at the effect of transfer level and scenario on trustees' back transfer decisions. We proceed by identifying the subjects who elicit reciprocal concerns. Then, we analyze reciprocators' back transfers across scenarios as well as their scenario choices in order to test how reciprocators behave when there is moral wiggle room. This is followed by a discussion of our results in the light of the related literature.

#### **3.1 Analysis**

We first perform a random-effects panel regression with the back transfer as the dependent variable. The panel includes all choices of a trustee (five transfer levels in three/four different scenarios). Standard errors are robust and clustered at the individual level. See table 1 for results. The regression includes the transfer received and dummies for scenarios 2 to 4. Overall, there is a positive correlation between the trustor's transfer and the amount trustees chose to return. On average, subjects seem to reciprocate. All scenario dummies are negatively correlated with the back transfer. On average, subjects seem to reduce the amount they send back when the scenario implies uncertainty about the implementation of their back transfer. Control variables (age, gender) are not statistically significant. This aggregate level result is in line with hypothesis 1.

We continue the analysis at the individual level. Following Fischbacher et al. (2001) we categorize subjects based on what they return (given a transfer of 0, 2.5, 5, 7.5 or 10) when they make a choice under certainty (scenario 1), see also table 2. Eleven subjects

Table 1: Determinants of amount returned

	coefficient	standard error
Transfer	1.224	.0633 ***
Scenario 2	-.245	.0934 ***
Scenario 3	-.4003	.1299 ***
Scenario 4	-.6691	.2416 ***
Age	.205	.1603
Female	1.037	.7636
Constant	-4.653	3.608
$R^2$		0.47

Panel regression with robust standard errors; 2,240 observations;

significance levels: \*\*\* = 1%, \*\* = 5%, \* = 10%

do not return anything, ever. The back transfers of 109 subjects are increasing weakly monotonous with the amount received and they are classified as conditional cooperators. Eight subjects elicit a humpback-shaped back transfer pattern. They first increase their back transfers with the amount received, but then decrease them. Our analysis considers conditional cooperators (even if they return only very little) as well as only partially reciprocating subjects (humpback-shaped pattern) as reciprocators. Along the way of our analysis we will test for the robustness of our results if selfish and humpback-shaped reciprocators are excluded.

Table 2: Categorization of subjects' scenario 1 back transfers

Type	Number of subjects	Mean of returned amount when receiving				
		0	2.5	5	7.5	10
Purely selfish	11	0	0	0	0	0
Conditional cooperators	109	.27	4.02	7.53	10.99	14.8
Humpback-shaped	8	0	4	6.37	7.87	3.87

What is reciprocating trustees' *behavior across scenarios*? More specifically, how did they behave in scenarios 2 to 4, that is, when there is a positive probability that their back transfer could fail? Table 3 reports the number of reciprocating subjects who returned less/same/more in the uncertain scenario (compared to scenario 1) for each amount received. We perform Wilcoxon signed-rank tests for each transfer level of

scenarios 2 to 4 in order to compare reciprocating subjects' choices under uncertainty to their scenario 1 choices. Significance levels are indicated in table 3. The majority of subjects does not change the back transfer, yet there is a general tendency to return less under uncertainty. For a relatively high chance of transfer success (90%, scenario 2) the tendency to decrease the back transfer is only significant for amounts received of 2.5 or 5. For an 80% chance of transfer success (scenario 3) the tendency to decrease the back transfer is significant for all amounts received except 0. In scenario 4 (implementation probability 50%, 57 reciprocators) the proportion of subjects who decrease is significant for all amounts received.

Table 3: Pairwise comparison of reciprocating subjects' back transfers

amount received	0	2.5	5	7.5	10
scenario 2 (90%)	7/106/4	19/91/7 **	19/90/8 **	20/84/13	16/92/9
scenario 3 (80%)	9/105/3 *	29/79/9 ***	32/71/14 ***	32/69/16 **	27/77/13 **
scenario 4 (50%)	4/53/0 **	17/33/7 **	19/32/6 ***	18/31/8 **	18/33/6 ***

We proceed to categorize subjects based on their choices across scenarios. For this purpose we compute, for every transfer level, the difference between back transfers in scenario 1 and 2, 2 and 3, and, if applicable, 3 and 4. The sum of these partial differences expresses how a subject reacted to the variation of the transfer implementation probability. We distinguish between three different behavioral patterns. Some trustees decreased their back transfers with the likelihood that the transfers fails. For each transfer level some trustees returned the same amount independently of the scenario. Finally, some increased their back transfers the more probable it gets that their transfer does not get implemented. Table 4 provides frequencies of these behavioral patterns. The categories appear to be similarly represented in sessions with three and four scenarios. A ranksum test ( $p = .88$ ) does not reject that the distribution of types is the same. Out of 128 subjects (all sessions pooled) 11 never return anything, 41 decreased, 55 did not change and 21 increased the back transfer across scenarios.<sup>6</sup> Table 4 also reports the

<sup>6</sup>The categorization aggregates over choices at all five transfer levels. Hence, it could be that a subject's behavior is inconsistent across transfer levels. One out of 55 subjects categorized as returning the same amount did in fact decrease the back transfers by 2.5 at a transfer level of 2.5 and increased them by 2.5 when receiving 5. All others never deviated from their scenario 1 back transfers. Among subjects categorized as increasing the amount two slightly decreased their back transfer at a transfer level of 0. All others never lowered the back transfer. Out of 41 subjects categorized as decreasing the

mean aggregate back transfers in scenario 1 of each category, that is, the sum of the five back transfer choices. Aggregate back transfers under certainty are not significantly different across categories.

Table 4: Categorization based on the back transfer schedules across scenarios

	total subjects	back transfers always 0	decreased amount across scenarios	returned same amount in all scenarios	increased amount across scenarios
3 scenarios	64	4	21	30	9
4 scenarios	64	7	20	25	12
all	128	11	41	55	21
mean (st. error) of aggregate back transfers in scenario 1	128	0	36.29 (2.61)	36.18 (2.08)	38.05 (3.71)

Reciprocating trustees' behavior across scenarios indicates that 41 subjects exploited moral wiggle room by reducing their back transfers with the likelihood that the transfers fail. Did these subjects tend to return zero with a positive failure probability or did they tend to exploit the moral wiggle room in a more subtle way? Overall, the majority seems to return only slightly less, although few subjects drop their back transfer to zero in uncertain scenarios. Figure 2 shows histograms of back transfers for scenarios 1 to 3 given a transfer of 10. It serves to illustrate the behavioral pattern among subjects who decreased their back transfers. The number of subjects returning zero increases when the failure chance of the back transfers is positive, but mostly subjects decide to return less.

The way reciprocators handle the variation of the back transfer success rate across scenarios has implications for our *analysis of the scenario choice*. In a strict sense, being able to pick a scenario only creates a situational excuse for subjects who returned the very same positive amount independently of the scenario. A subject who increased transfers across scenarios may have done so to make sure that, in (subjective) expectations, the same amount reaches the trustor. In such a case, being in a position to select a scenario with an implementation probability less than 1 would not be advantageous for the subject's expected utility.<sup>7</sup> Finally, subjects who decreased amounts may back transfers one subject increased the amount returned at a transfer level of 5 and one subject was inconsistent. All others never increased the back transfer. If selfish and humpback-shaped reciprocators are considered, 21 subjects are categorized as selfish, 33 decrease, 47 return the same and 19 increase the amount.

<sup>7</sup>Note that we do not require subjects to comply with expected utility theory as we cannot assume

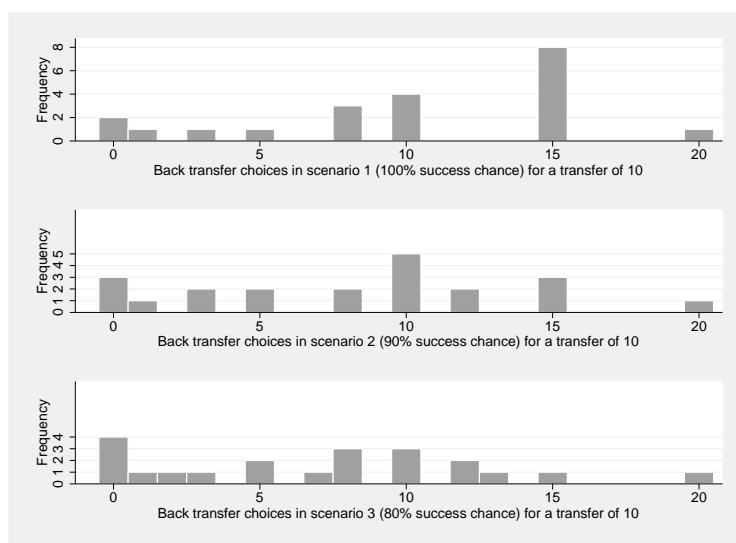


Figure 2: Histograms of back transfers for scenarios 1 to 3 at a transfer of 10

have already exploited moral wiggle room when they made their back transfer choices in scenarios with uncertainty. They would only benefit from these choices by actually picking a scenario with an implementation probability less than 1. It is not clear how they would react to a second serving of moral wiggle room, though.<sup>8</sup> Hence, our test of the moral wiggle room effect at the scenario choice focuses on subjects who did not vary the back transfers across scenarios.

Figure 3 shows histograms of the scenario choice for the four categories: purely self-ish, decreasing, same, and increasing back transfers across scenarios. For trustees who returned the same amounts across scenarios (figure 3, bottom left) the scenario choice involved the unambiguous opportunity to reap a monetary gain (in expectations). In this category 33 of 55 subjects selected scenario 1. In contrast, 22 of them made use of the moral wiggle room and picked a scenario that did not guarantee the back transfer. A one-sided binomial test confirms that this fraction is significantly greater than zero ( $p < 0.01$ ) and supports hypothesis 2. Subjects who increased back transfers across scenarios (figure 3, bottom right) likely have equalized their subjective expected utility subjects calculate this properly. The sequence of choices needed to look like they tried to equalize the actual amount. This means weakly monotonically increasing amounts returned.

<sup>8</sup>They may consciously choose the scenario that maximizes their expected payoff. However, having to pick a scenario that clearly favors them may be too much to still appear pro-social. Then, a choice of a less favorable scenario would result. Their choice might also be affected by moral balancing keeping them from exploiting moral wiggle room two times in a row.

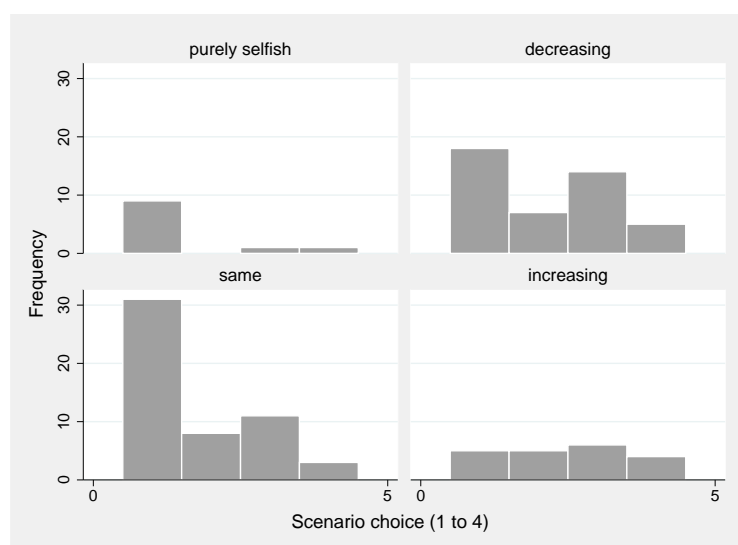


Figure 3: Scenario choices by categories

across scenarios. Then, their choice of the scenario should not matter to them and it should be uniformly distributed. This seems to be the case. Subjects who decreased amounts (figure 3, top right) appear to have already exploited moral wiggle room when they made their back transfer choices in scenarios 2 to 4. No clear pattern with respect to their scenario choice seems evident. Finally, the scenario choice of purely selfish subjects (figure 3, top left) has no consequence for their payoffs.

### 3.2 Discussion of our results

In our experiment, 41 reciprocating subjects decreased their back transfer when the failure chance of the transfer was positive, an indication that they made use of this situational excuse. However, 16 of them eventually made a scenario choice that is clearly disadvantageous to them, while 25 selected a scenario that favors their expected payoffs. Out of 55 reciprocators who returned the same positive amount independently of the situation 22 selected a scenario that implied a positive chance that the back transfer fails to reach the trustor. The remaining 33 selected scenario 1 and made sure the back transfer reaches the trustor. They made no use of the moral wiggle room in the scenario 2 to 4 back transfer choices and resisted the moral wiggle room provided by the scenario choice. Finally, 21 reciprocators increased the back transfer across scenarios, thus, resisting our first and evading our second manipulation. Summarizing, 47 of 117 (40%) reciprocators exploited moral wiggle room, while 70 (60%) resisted (to some



extent).<sup>9</sup> Our study also offers some insight about the way subjects exploited moral wiggle room. When they drop their back transfer to zero in uncertain scenarios, their decisions seem to be influenced by a desire not to appear selfish to others. When they return only slightly less in uncertain scenarios, their decisions seem to be influenced by a desire not to appear selfish to themselves. We find evidence for both types of exploiting moral wiggle room. Finally, it is worth to note that two of our design choices made our experiment a tougher test environment for moral wiggle room to prevail than comparable experiments. While the side-by-side interface for entering back transfer schedules makes it easier for subjects who would like to enter the same positive amounts across scenarios to do so, it may become more difficult for subjects who have a tendency to exploit moral wiggle room to actually do so. Since scenario 1 choices are still visible, the context of the choices under uncertainty is more salient than without the reminder. Moreover, we let our subjects play both roles which means that trustees are familiar with the trustor's perspective of the situation. This potential awareness about the other role may make it harder to exploit moral wiggle room in comparison to a design in which subjects only play one role.

### **3.3 Comparison to the related literature**

Similar to previous studies in dictator game contexts we find that a substantial amount of subjects makes use of moral wiggle room. Most of these related studies are between-subjects designs which complicates a quantitative comparison of the results. Our within-subjects design endogenizes the decision to reciprocate allowing us to identify selfish subjects (11 of 128 (8.6%)) and measure the tendency to exploit wiggle room among actual reciprocators (47 of 117 (40%)). In related<sup>10</sup> between-subjects designs, the level of selfish choices observed in the baseline could be regarded as an estimate for the selfish type and the increase of selfish choices from baseline to treatment as an estimate for pro-social subjects who exploit moral wiggle room. Applying this to data from Dana et al. (2007)<sup>11</sup> suggests 26.3% selfish types and a range of 36.2% to 39.2% for 'wiggling'

---

<sup>9</sup>When excluding selfish subjects and humpback-shaped reciprocators, 60 out of 99 (61%) resisted.

<sup>10</sup>In this comparison we focus on one-shot studies that employ a standard mini dictator game as baseline and a moral wiggle room manipulation as treatment.

<sup>11</sup>Dana et al. (2007) employ three variations of 'moral wiggle room' and compare them to a baseline mini dictator game. In the 'multiple dictator' treatment, a second dictator was added. Either dictator

pro-socials. Larson and Capra (2009) run the ‘hidden information’ set up in a double-blind environment. They report 22% selfish choices in their baseline and 56% in the treatment. Andreoni and Bernheim (2009) conduct a dictator game with the possibility of nature forcing a choice of 0. In the baseline (when the probability of a forced choice is 0) the rate of selfish choices is 30% and in the treatments with a high probability of a forced choice (e.g. 50% or 75%) it is around 70%. Grossman (2014) adds variations to the ‘hidden information’ design. In his baseline 34.6% take the selfish choice and results from the original treatment suggest a rate of 24.9% for subjects who exploit wiggle room. Finally, Matthey and Regner (2015) employ a similar manipulation of moral wiggle room as our study (the only difference besides the dictator/trust setting is that the implementation probability in their three uncertain scenarios varies between 0.8 and 0.9). Hence, their within-subjects design endogenizes the decision to send, identifying selfish types (17.8%) and pro-socials who make use of situational excuses (38.1%). Table 5 summarizes the comparison.

Table 5: Overview of related studies and their findings

study	game type	study design	manipulation used	estimate of pro-selfs	estimate of ‘wigglers’
Dana et al. (2007)	DG	between	MD HI PD	26.3% 26.3% 26.3%	38.7% 36.2% 39.2%
Larson and Capra (2009)	DG	between	HI	22%	56%
Andreoni and Bernheim (2009)	DG	between	FC	30%	40%
Grossman (2014)	DG	between	HI	34.6%	24.9%
Matthey and Regner (2015)	DG	within	U	17.8%	38.1%
van der Weele et al. (2014)	TG	between	PD	62.5%	–
this study	TG	within	U	8.6%	36.7%

Notes: Game type: dictator game (DG) or trust game (TG); manipulation used: ‘multiple dictator’ (MD), ‘hidden information’ (HI), ‘plausible deniability’ (PD), ‘forced choice’ (FC) or uncertainty (U); estimates of pro-selfs and ‘wigglers’ are provided as percentages of the total samples.

Also van der Weele et al. (2014) aim to test moral wiggle room in the context of reciprocity.<sup>12</sup> They hypothesize that reciprocal behavior in the trust game is less manipulable than dictator game giving. In their between-subjects design trustors are endowed with 20 Euro and can send either 0, 10 or 20 Euro. The transfer is tripled and trustees could implement equal payoffs for all. In the ‘plausible deniability’ treatment, the dictator was cut off if a decision was not made fast enough. Then, the computer picked either outcome with equal probability. In the ‘hidden information’ treatment, dictators could remain ignorant about the precise consequences of their choice to the recipient.

<sup>12</sup>The ideas for these two papers were developed independently of each other.

face a binary choice (return nothing or two thirds), yielding payoffs of (10, 50) or (0, 80) or an equal split of the pie (30, 30 or 40, 40). They adapt the ‘plausible deniability’ treatment of Dana et al. (2007), that is, in the treatment the trustee might get cut off by the computer. In this case the two options of the trustee (return nothing or two thirds) are implemented with equal probability. The trustor will not find out whether the trustee or the computer made the decision. Van der Weele et al. (2014) do not find a treatment difference and argue that providing moral wiggle room has no effect on the incidence of reciprocal behavior. Our findings do not contradict their hypothesis: reciprocal behavior may well be less prone to moral wiggle room than giving in a dictator game context. However, our findings suggest a qualification of their conclusion. In contrast to their study we find a significant effect of moral wiggle room on reciprocal behavior.

What may be an explanation for the discrepancy between their results and ours? One possibility could be that in our design the trustee’s choice is continuous, not binary as in van der Weele et al. (2014). Hence, our design would identify a substantial tendency to exploit moral wiggle room, even if the behavior change due to the manipulation is relatively subtle. In contrast, a binary choice set might not pick up a substantial tendency to exploit wiggle room, if the tendency is not large enough to switch behavior from splitting equally to returning nothing. Another potential source for the discrepancy could be that one study uses a within-subjects, the other a between-subjects design.

A look across the between-subjects moral wiggle room studies described in table 5 points at a further explanation. The fraction of selfish choices in the baseline is never more than 35%, while it is 62.5% in van der Weele et al. (2014). The fraction of selfish choices in the treatment (i.e., selfish and ‘wiggling’ subjects combined) is between 59% and 78% in other studies. Essentially, a fraction of subjects seems to be immune to the employed moral wiggle room manipulations. They would behave pro-socially in any case. Assuming that in their experiment there is a similarly sized fraction of subjects who would never make a selfish choice as in related studies, it seems that there is no scope for the moral wiggle room manipulation in their design. Hence, a sort of ceiling effect, due to the high rate of selfish behavior already in the baseline, may have limited the effect of the treatment manipulation.

To summarize, in most of the surveyed dictator game studies the fraction of ‘wiggling’ pro-socials hovers around 40% (the 56% in the double-blind study of Larson and Capra (2009) being an exception). The tendency to exploit moral wiggle room in our study’s trust game context is at a similar level. The fraction of selfish types (8%) is, however, substantially lower than in the dictator game studies. Of course, this is only a comparison across different studies and partly across different designs/manipulations. It should be an interesting path for future research to establish whether moral wiggle room affects reciprocal concerns really to the same extent as the preference to give.

Finally, in a stream of literature that is related to the moral wiggle room dictator games, Dana et al. (2006), Broberg et al. (2007) and Lazear et al. (2012) analyze subjects’ behavior when an ‘exit option’ to get out of a dictator game is provided. A substantial amount of subjects avoids the dictator game, even if they get a lower payoff. Malmendier et al. (2014) introduce the ‘exit option’ into a double dictator game in order to test how reciprocity affects avoidance. They find a substantial level of sorting out in a positive reciprocity condition (about 30%) although avoidance is higher in a neutral condition (50%).

## **4 Conclusion**

We conducted a modified trust game in order to analyze how reciprocators respond to the provision of moral wiggle room. In our experiment a substantial amount (40%) of reciprocating subjects behaved less pro-social when we introduced moral excuses for selfish behavior. That is, when the context of their choice became less salient, they succumbed to the temptation of keeping more.

Our evidence shows that moral wiggle room effects extend beyond the setting of a dictator game where they have been established so far to the one of a trust game. It seems that the preference to reciprocate is also affected by the availability of situational excuses, just as the preference to give.

## References

- Akerlof, G. A. and R. E. Kranton (2000). Economics and Identity. *Quarterly Journal of Economics*, **115**, pp 715–753
- Akerlof, G. A. and R. E. Kranton (2005). Identity and the economics of organizations. *Journal of Economic Perspectives*, **19**, pp 9–32
- Andreoni, J. and B. D. Bernheim (2009). Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica*, **77**(5), 1607–1636
- Aronson, E. (1992). The return of the repressed: Dissonance theory makes a comeback. *Psychological Inquiry*, **3**, 301–311
- Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, **11**, 122–133.
- Beauvois, J.-L. and R.-V. Joule (1996). A Radical Theory of Dissonance. European Monographs in Social Psychology. *Taylor and Francis*, New York, NY
- Bénabou, R. and J. Tirole (2006). Incentives and Pro-Social Behavior. *American Economic Review*, **96**(5), 1652–1678
- Bénabou, R. and J. Tirole (2011). Identity, Morals and Taboos: Beliefs as Assets. *Quarterly Journal of Economics*, **126**(2), 805–855
- Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, Reciprocity and Social History. *Games and Economic Behavior*, **10**, 122–142.
- Broberg, T., Ellingsen, T., and Johannesson, M. (2007). Is generosity involuntary? *Economics Letters*, **94**, 32–37.
- Cherry, T. L., Frykblom, P., and J. F. Shogren (2010). Hardnose the dictator. *American Economic Review*, **92**(4), 1218–1221.
- Conrads, J. and B. Irlenbusch (2013). Strategic ignorance in ultimatum bargaining. *Journal of Economic Behavior & Organization*, **92**, 104–115.

- Dana, J., Cain, D.M., and Dawes, R.M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, **100**, 193–201.
- Dana, J., Weber, R., and Kuang, J.X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, **33**, 67–80.
- Festinger, L. (1957) A Theory of Cognitive Dissonance. *Row Petersen*, Evanston, IL
- Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics*, **10**(2), 171–178.
- Franzen, A. and S. Pointner (2012). Anonymity in the dictator game revisited. *Journal of Economic Behavior & Organization*, **81**, 74–81.
- Greiner, B. (2004). An Online Recruitment System for Economic Experiments. *University of Cologne working paper series*
- Grossman, Z. (2014). Strategic Ignorance and the Robustness of Social Preferences. *Management Science*, **60**(11), 2659–2665.
- Grossman, Z. (2015). Self-signaling and social-signaling in giving. *Journal of Economic Behavior & Organization*, forthcoming.
- Guala, F. and L. Mittone (2010). Paradigmatic experiments: The Dictator Game. *Journal of Socio-Economics*, **39**, 578–584.
- Haisley, E. C. and R. A. Weber (2010). Self-serving interpretations of ambiguity in other-regarding behavior. *Games and Economic Behavior*, **68**, 614–625.
- Hamman, J. R., Loewenstein, G., and R. A. Weber (2010). Self-Interest through Delegation: An Additional Rationale for the Principal-Agent Relationship. *American Economic Review*, **100**, 1826–1846.
- Hoffman, E., McCabe, K., and V. L. Smith (2010). Social distance and other-regarding behavior in dictator games. *American Economic Review*, **86**(3), 653–660.
- Konow, J. (2000). Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review*, **90**(4), 1072–1091.

- Larson, T. and C. M. Capra (2009). Exploiting moral wiggle room: Illusory preference for fairness? A comment. *Judgment and Decision Making*, **4**, 467–474.
- Lazear, E., U. Malmendier and R. A. Weber (2012). Sorting in Experiments with Application to Social Preferences. *American Economic Journal: Applied Economics*, **4**, 136–163.
- List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, **115**(3), 482–494.
- Malmendier, U., V. L. te Velde and R. A. Weber (2014). Rethinking Reciprocity. *Annual Review of Economics*, **6**, 849–874.
- Matthey, A. and T. Regner (2011). Do I really want to know? A cognitive dissonance-based explanation of other-regarding behavior. *Games*, **2**, 114–135.
- Matthey, A. and T. Regner (2015). More than outcomes: The role of self-image in other-regarding behavior. *Review of Behavioral Economics*, forthcoming
- Mazar, N., O. Amir and D. Ariely (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, **XLV**, 633–644.
- van der Wee, J. (2013). Inconvenient Truths: Determinants of Strategic Ignorance in Moral Dilemmas. *working paper*
- van der Wee, J., J. Kulisa, M. Kosfeld and G. Friebe (2014). Resisting Moral Wiggle Room: How Robust is Reciprocal Behavior? *American Economic Journal: Microeconomics*, **6**, 256–264.
- Trautmann, S. and F. Vieider (2012). Social Influences on Risk Attitudes: Applications in Economics. *Handbook of risk theory*, eds: Roeser, S., Hillerbrand, R., Sandin, P., and M. Peterson

## Instructions

**Welcome and thank you for participating!** In this experiment you can earn a monetary amount depending on your decisions and the decisions of the other participants. **Therefore, it is very important that you read the instructions carefully.**

Please note that these instructions are directed to you only and you are not allowed to exchange any information with the other participants.

Also, it is not allowed to talk to other participants during the whole experiment. Whenever you have a question please raise your hand. We will come to your place and answer your question. Please never ask your question(s) aloud. In case you break these rules we will have to end the experiment. Please switch off your mobile phones now.

### General procedure

The experiment will take around 60 minutes. Your earnings from this experiment depend on your decisions and possibly on the other participants' decisions. All amounts in the decision situations are stated **in Euro**. The exact amount will be paid to you at the end of the experiment. Additionally, you will receive 2.50 Euro for your participation in the experiment.

After filling out a questionnaire the experiment will be finished and you will receive your payment.

Overview of the procedure:

- reading the instructions, answering the control questions
- decision situations
- questionnaire
- payment and end of the experiment

### Details

In the experiment you will be randomly paired with another participant. This means two participants (referred to as A and B) will interact. In the course of the experiment you will take decisions in the role A as well as in the role B. Which role (A or B) will be relevant for your payout will be decided randomly after the experiment. Therefore, it is very important that you **familiarize with both roles**.

Both participants receive an *endowment* of 10 Euro.

First, A will make a decision. He/she can send an amount between 0 and the *endowment* of 10 Euro to participant B. This *transfer* will be **tripled** and given to B.

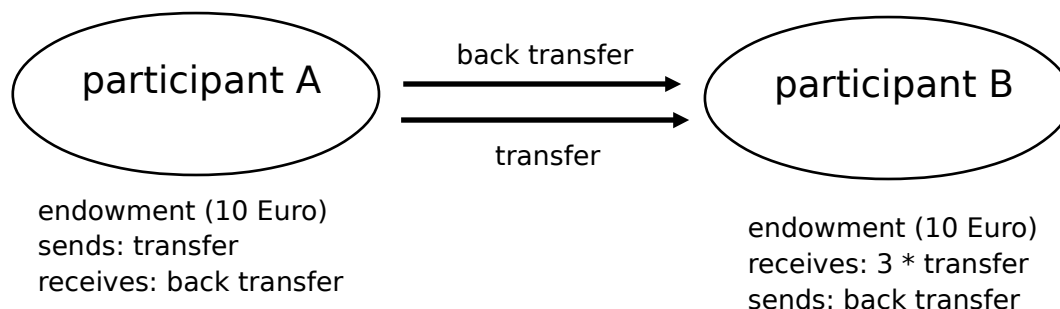
After that, participant B can decide how much he/she wants to send back to participant A. This *back transfer* can vary between 0 and 10 Euro + 3 \* *transfer*.

The following payoffs will result accordingly:

- participant A:  $endowment - transfer + back\ transfer$
- participant B:  $endowment + 3 * transfer - back\ transfer$



The following diagram illustrates the game and the resulting payoffs:



There are 4 different situations for the implementation of the game:

*Situation 1:* Here, the *back transfer* will be transferred with certainty; this means the *back transfer* will definitely be subtracted from B's account and then added to A's account.

*Situation 2:* Here, the *back transfer* will be transferred with a 90% chance; thus, the chance of it not being transferred is 10%.

*Situation 3:* Here, the *back transfer* will be transferred with an 80% chance; thus, the chance of it not being transferred is 20%.

*Situation 4:* Here, the *back transfer* will be transferred with a 50% chance; thus, the chance of it not being transferred is 50%.

In situations 2 to 4 the following applies:

**In case the *back transfer* is transferred**, the amount will be subtracted from B's account and then added to A's account. The resulting payoffs are as shown above:

- participant A:  $endowment - transfer + back\ transfer$
- participant B:  $endowment + 3 * transfer - back\ transfer$

**In case the *back transfer* is not transferred**, the accounts of A and B will remain unchanged; this means the *back transfer* won't have any impact on the payoffs:

- participant A:  $endowment - transfer$
- participant B:  $endowment + 3 * transfer$

Procedure:

1. Entry of the *transfer* which as participant A you want to send to B. The *transfer* can be chosen from the following values: 0 Euro, 2,5 Euro, 5 Euro, 7,5 Euro, 10 Euro.
2. Entry of the *back transfer*, which as participant B you want to send back to A. You will enter the *back transfer* (in multiples of 10 Cents) for each possible *transfer* (0; 2,5; 5; 7,5; 10). Additionally, you will enter *back transfers* for each of the 4 situations described above.

3. Then, it will be decided which situation occurs. In case situation 2, 3 or 4 is chosen the computer will calculate according to the respective probabilities whether the *back transfers* will be transferred. Afterwards, if applicable the *back transfers* will be transferred.

## **Payoff**

Which role (A or B) will be the relevant role for your final payoff will be decided by chance (50% / 50%). Accordingly, there are two possibilities:

1. You receive your payoff in the role of participant A. In case situation 1 has been chosen your payoff equals *endowment – transfer + back transfer*. In case situation 2, 3 or 4 has been chosen your payoff equals *endowment – transfer + back transfer* if the *back transfer* is indeed transferred. If it is not transferred your payoff equals *endowment – transfer*.
2. You receive your payoff in the role of participant B. In case situation 1 has been chosen your payoff equals *endowment + 3 \* transfer – back transfer*. In case situation 2, 3 or 4 has been chosen your payoff equals *endowment + 3 \* transfer – back transfer* if the *back transfer* is indeed transferred. If it is not transferred your payoff equals *endowment + 3 \* transfer*.