Actions and the self: I give, therefore I am?

Tobias Regner [♡*]Astrid Matthey [♠]

[♠]Max Planck Institute of Economics, Jena, Germany

[♡]*University of Jena, Jena, Germany*

Abstract

Self-signaling models predict less selfish behavior in a probabilistic giving setting as individuals are expected to invest in a pro-social identity. However, there is also substantial evidence that people tend to exploit situational excuses for selfish choices (for instance, uncertainty) and behave more selfishly. We contrast these two motivations experimentally in order to test which one is more prevalent in a reciprocal giving setting. 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 when self-image relevant factors are varied. Our results indicate that self-deception is prevalent when subjects make the back transfer choice. Twice as many subjects seem to exploit situational excuses than subjects who appear to invest in a pro-social identity.

JEL classifications: C72, C91, D80, D91

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

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 can be kept? Dana, Weber and Kuang (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, Loewenstein and Weber, 2010; Matthey and Regner, 2011; Feiler, 2014; Grossman, 2014; van der Weele, 2014; Exley, 2015) confirms such a self-serving bias in dictator game giving.

Bayesian self-signaling models propose a different type of behavior in an allocation decision: individuals are unsure whether they are a pro-self or pro-social type, derive utility the more they believe they are a pro-social type and, thus, may give in order to send a positive signal to their self. Grossman (2015) develops such a model and tests it experimentally in a binary probabilistic dictator game. The model predicts more giving in the low probability treatment as the expected cost of sending a pro-social signal is cheapened. However, results do not lend supporting evidence.[1]

Bénabou and Tirole (2011) consider further channels how actions and the self relate to each other. Besides identity management (beliefs about one's type are malleable through actions) they also allow, for instance, for the possibility that an action may be regarded as uninformative about the self. Hence, individuals may attribute their selfish action to the context, instead of having to connect selfish behavior to their self-image. The salience of the context – to what extent one's action has an effect on the outcome – systematically varies the informativeness of an action. As a consequence, individuals would have a higher tendency to invest in their identity, when informativeness is high. Likewise, they would tend to succumb to the temptation of situational excuses, when informativeness is low. Naturally, it is in the self's 'eye of the beholder', whether an action is perceived to convey information about one's underlying character (potentially leading to an identity investment) or whether situational excuses are invoked and used to bias the signal to the self (making the identity damage of a selfish choice acceptably small). Thus, it is an empirical question which motivation is more dominant. Our study's goal

---

[1] Also van der Weele and von Siemens (2014) test Bayesian self-signaling experimentally but find no empirical support. Grossman and van der Weele (2017) develop a Bayesian self-signaling model in the context of information acquisition where individuals can willfully ignore to get informed. Their related experimental test supports the model.

is to consider both self-signaling channels (identity management and self-deception) and to test them experimentally.

We set up a probabilistic giving environment in which the predictions of the two approaches contrast each other. Identity management predicts higher transfers with increased uncertainty about the implementation of the transfer due to the pro-social signal becoming cheaper. Instead, self-deception predicts lower transfers, because uncertainty about the actual implementation could serve as a situational excuse.

Since the effect of moral wiggle room on giving in dictator games is well established, we decided to move our test bed to a less explored domain, namely, when also reciprocal concerns may motivate individuals.[2] 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 behave. 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 excuse to return less in these scenarios (alternatively, the decreased implementation chance of the transfer could induce subjects to return more via identity management). 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.

Our within-subjects design allows us to analyze back transfer choices at the individual level. Thus, we can distinguish between trustees motivated by self-deception and identity management. While our results show that self-deception is more common when subjects make the back transfer choice (twice as many subjects decrease than increase their transfers under uncertainty), they also indicate that both motivational processes appear relevant for human decision making. Furthermore, as a substantial fraction of subjects makes a self-serving scenario choice, our

---

[2]Results of existing studies which analyze reciprocal behavior when there is a situational excuse for not giving (van der Weele et al., 2014; Regner, in press) point in different directions. Thus, our study will provide an additional data point.

results indicate that reciprocators make use of moral wiggle room if situational excuses exist.

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, Dickhaut and McCabe, 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 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 3 (a) (in the appendix) 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. 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, 2004; 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 self-image relevant factors are varied. We use the model of Bénabou and Tirole (2011) to guide our analysis.[3] A key component of it is that beliefs about one's pro-sociality type are malleable through actions as imperfect recall is assumed. Thus, identity management becomes possible: if the cost of sending a pro-social signal (via performing a pro-social action) is small enough, pro-self types

---

[3]Several approaches exist to model self-image concerns. See also the literature on cognitive dissonance (Festinger, 1962; Aronson, 1992; Beauvois and Joule, 1996; Konow, 2000; Spiekermann and Weiss, 2016), identity (Akerlof and Kranton, 2000), self-impression management (Murnighan, Oesch and Pillutla, 2001), self-concept maintenance (Mazar, Amir and Ariely, 2008) and other self-signaling models (Bodner and Prelec, 2003; Bénabou and Tirole, 2006; Tirole, Falk and Bénabou, 2016).

decide to invest in their identity by imitating a pro-social type.[4] Bénabou and Tirole (2011) also allow for the possibility that inferences from actions about one's type are malleable. Such inferential wiggle room exists, if the informativeness of an action about one's self-image is imperfect.[5] Inferential wiggle room allows self-deception: if the salience of a situation is low enough (reducing the signal strength of an action), pro-social types tend to make a selfish choice as the monetary gain outweighs the negative effect on the self-image.

Our experimental design manipulates self-image related aspects at two stages. First, trustees could exploit the fact that the back transfer is not executed for sure in scenarios 2 to 4.[6] Second, trustees could succumb to the temptation of choosing a scenario that benefits them.

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 situations in scenarios 2 to 4 allow 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. Haisley and Weber (2010) find such a self-serving bias caused by uncertainty in a related study involving dictator game choices under ambiguity. See also Exley (2015) who studies choices between a certain amount and risky lotteries. She varies the recipient of both (self vs. a charity) and finds evidence of the use of risk as an excuse to give less. See Shalvi et al. (2015) for an overview of self-serving justifications used in the related domain of ethical/moral behavior. Thus, we expect that some trustees engage in self-deception when they make the back transfer choice. Moreover, the effect of uncertainty should be more pronounced the lower the implementation probability of the back transfer is. Thus, we expect a positive relationship between the probability and their back transfer choice.

Hypothesis 1 *In comparison to scenario 1 a non-trivial fraction of reciprocating trustees trans-*

---

[4]Grossman (2015) focuses on this aspect of self-signaling. He tests his model experimentally in a binary probabilistic dictator game. The transfer's implementation probability is either certain or 1/3. The model predicts that giving in the low probability treatment is higher as the expected outcome-utility cost of the pro-social choice is cheapened. However, his results do not offer conclusive supporting evidence.

[5]See their parameter ν and their footnote 10. See also Tirole, Falk and Bénabou (2016) who endogenize the signal strength of an action by incorporating narratives in their self-signaling model. Such a narrative, provided by others or 'self-authored', may serve as an excuse for immoral behavior. It allows individuals to rationalize an immoral action even though it only contains a minimal degree of subjective plausibility.

[6]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.

fer back less when the transfer could fail (scenarios 2 to 4).

Bayesian self-signaling would predict the opposite effect on trustees' back transfer choices. A decrease of the transfer's implementation probability cheapens the pro-social signal to the self. Investing in identity becomes more affordable and, in turn, more pro-social choices should result. If trustees engage in identity management, we expect a positive relationship between the probability and their back transfer choice.[7]

Hypothesis 2 *In comparison to scenario 1 a non-trivial fraction of reciprocating trustees transfer back more when the transfer could fail (scenarios 2 to 4).*

A second situational excuse 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 3 *When the choice of a scenario that involves uncertainty results in an expected monetary gain, a substantial amount of reciprocating trustees does not choose scenario 1.*

Finally, 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) and is also in line with the prediction for the "Probability & Outcome" treatment in Grossman (2015).[8] 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

---

[7]Note that an alternative motivation to increase transfers could be the desire to make sure that the same amount reaches the trustor, independently of the scenario.

[8]Note, however, that in our design we cannot be completely sure that trustees assume the trustor will be informed about the possibility that the transfer may fail when trustees make their back transfer choice. They may assume that trustors only learn the outcome. In that case, social-signaling does not predict an effect of a variation of the implementation probability on the incidence of zero back transfers. See Grossman (2015) for the formal argument.

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 (with respect to the scenario choice) 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

Using the ORSEE software (Greiner, 2004) 128 subjects were recruited among students from various disciplines at the local university. 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).

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. Subjects 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 they behave when self-image relevant factors are varied.

3.1 Analysis

We first perform random-effects panel regressions 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 specification in column I includes the transfer received and the implementation

probability as explanatory variables. Coefficients for both are positive and highly significant. The specification in column II replaces the implementation probability with dummies for scenarios 2 to 4. All scenario dummies are negatively correlated with the back transfer. Further tests of the dummy coefficients show that back transfers in scenario 3 and 4 are significantly lower than back transfers in scenario 2 ($p = .0549$, $p = .0664$). Back transfers in scenario 4 are not significantly lower than back transfers in scenario 3 ($p = .1507$). Overall, there is a positive correlation between the trustor's transfer and the amount trustees chose to return. On average, subjects reciprocate. Moreover, 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.

Table 1: Determinants of amount returned

	I:		II:	
	coefficient	standard error	coefficient	standard error
Transfer	1.224	.0632 ***	1.224	.0633 ***
Implementation probability	1.351	.4888 ***	–	–
Scenario 2 (90%)	–	–	-.245	.0934 ***
Scenario 3 (80%)	–	–	-.4003	.1299 ***
Scenario 4 (50%)	–	–	-.6691	.2416 ***
Age	.2053	.1603	.205	.1603
Female	1.035	.7637	1.037	.7636
Constant	-6.078	3.795	-4.653	3.608
Adjusted R^2	0.47		0.47	

Panel regression with robust standard errors; 2,240 observations; in II the scenario with certainty (1)

serves as the baseline; significance levels: $*** = 1\%, ** = 5\%, * = 10\%$

We continue the analysis at the individual level. Following Fischbacher, Gächter and Fehr (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 do not return anything, ever. The back transfers of 109 subjects are increasing weakly monotonically 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. At the end of our analysis we will test for the robustness of our results, if humpback-shaped and selfish 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 percentage of reciprocating subjects who returned less/same/more in scenarios 2 to 4 (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. 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 (at the 5%-level) 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 (at least at the 5%-level) for all amounts received except 0. In scenario 4 (implementation probability 50%) the proportion of subjects who decrease is significant (at least at the 5%-level) 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%)	6/90.6/3.4	16.2/77.8/6 **	16.2/76.9/6.9 **	17.1/71.8/11.1	13.7/78.6/7.7
Scenario 3 (80%)	7.7/89.7/2.6 *	24.8/67.5/7.7 ***	27.4/60.7/11.9 ***	27.4/60/13.6 **	23.1/65.8/11.1 **
Scenario 4 (50%)	7/93/0 **	29.8/57.9/12.3 **	33.3/56.2/10.5 ***	31.6/54.4/14 **	31.6/57.9/10.5 ***

Notes: In each cell x/y/z indicates the percentage of reciprocating subjects who returned less/same/more in the respective scenario in comparison to scenario 1. There are 117 reciprocators in scenarios 2 and 3, 57 in scenario 4. Significance of Wilcoxon signed-rank tests of choices under uncertainty compared to scenario 1 choices: * * * = 1%, ** = 5%, * = 10%

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.[9] Table 4 also reports 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. Moreover, reciprocators' transfer choices as trustor do not differ across categories (4.88, 4.73, 4.76), while the ones of purely selfish trustees are significantly lower (2.5).

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
aggregate back transfers in scenario 1, mean (st. error)	128	0	36.29 (2.61)	36.18 (2.08)	38.05 (3.71)
transfer choice as trustor, mean (st. error)	128	2.5 (1.21)	4.88 (0.33)	4.73 (0.41)	4.76 (0.29)

Reciprocating trustees' behavior across scenarios indicates that 41 subjects reduced 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 make use of the excuse 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 1 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. Under certainty, returning 15 (which corresponds to sending back half of what has been received) is the most popular choice of trustees. In scenarios 2 and 3, the number of trustees returning 15 drops sharply and smaller back transfers become more common. The number of subjects returning zero increases when the failure chance of the back transfers is positive, but mostly subjects decide to return less.

---

[9]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 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.
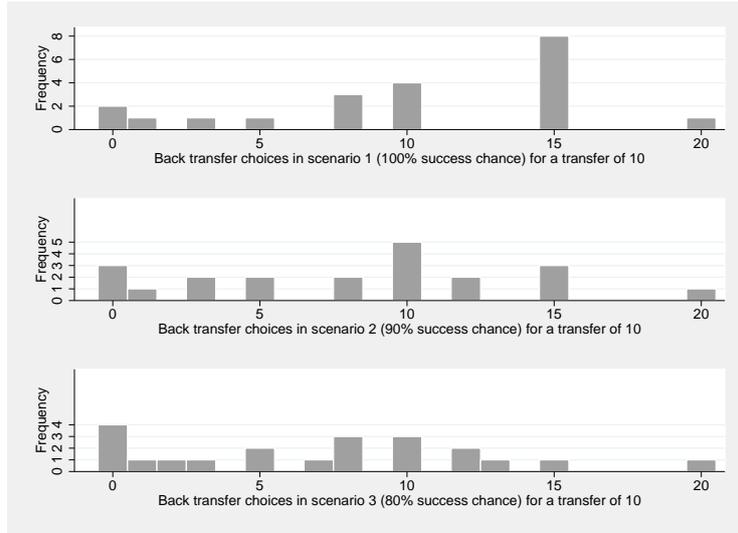
Figure 1: Histograms of back transfers for scenarios 1 to 3 (for subjects who decreased the amount and a transfer of 10)

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 opens up moral wiggle room only for subjects who returned the very same positive amount independently of the scenario. A subject who increased transfers across scenarios may have done so in order to invest in identity, thus conveying a signal of being pro-social. In such a case, being in a position to select a scenario with an implementation probability less than 1 may not be advantageous for the subject's expected utility. Finally, subjects who decreased amounts may 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.[10] Hence, our analysis of the scenario choice focuses on the 55 subjects who did not vary the back transfers across scenarios.

Figure 2 shows histograms of the scenario choice for the four categories: purely selfish, decreasing, same, and increasing back transfers across scenarios. For trustees who returned the same amounts across scenarios (figure 2, 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. Allowing for some noise in the decision

---

[10]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 engaging in self-deception two times in a row.
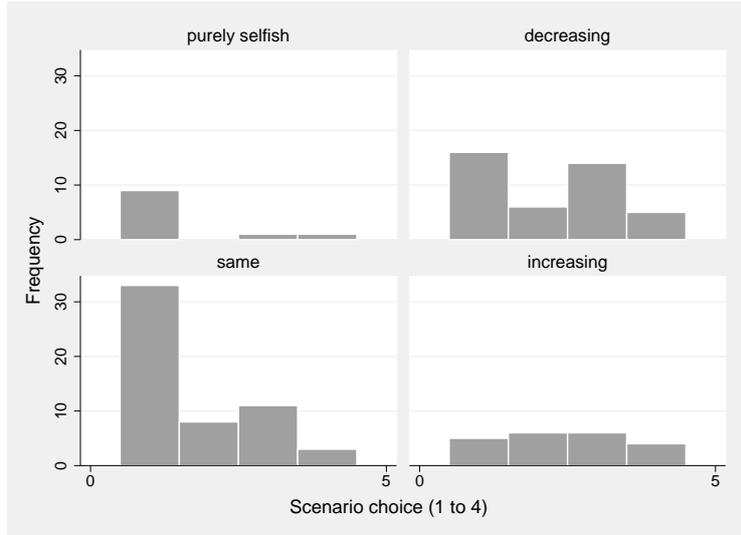
Figure 2: Scenario choices by categories

making (i.e. some trustees, say 25%, select a scenario other than 1 by chance), a one-sided binomial test confirms that this fraction is significantly greater than the noise level ($p = 0.01$) and supports hypothesis 2. When subjects increased back transfers across scenarios, their choice of the scenario should not matter to them and we may expect a uniform distribution. This seems to be the case (figure 2, bottom right). Subjects who decreased amounts (figure 2, 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 2, top left) has no consequence for their payoffs.

3.2 Discussion

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 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).[11]

The back transfer choice across scenarios may be affected by two self-signaling channels and in our experiment we find evidence for both behavioral patterns. However, self-deception is more common as 41 subjects seem to engage in it compared to 21 whose behavior is consistent with identity management.

If the transfer choice was binary, as in the model of Grossman (2015), identity management predicts low pro-sociality types take the pro-social choice given the cost of the signal cheapens sufficiently. High types are not expected to change their behavior. They already take the pro-social choice under certainty and cannot improve on that. Alternatively, uncertainty about implementation of the back transfer would trigger self-deception processes as the self would perceive the situation as an excuse to behave more selfishly. In a binary context, high types would engage in self-deception (if the psychological cost is small enough), while low types already take the selfish choice under certainty. This implies that in a binary setting the direction of the effect of p would depend on the prevalence of low/high pro-sociality types. By design, only low types can invest in identity and self-deception is exclusive to high types. Consequently, identifying either of the behavioral pattern requires sufficient low/high types in the role of the decision maker.

In our experiment, subjects have more than two transfer options to choose from leaving both types the theoretical possibility to go either way. Unless subjects choose an extreme in scenario 1, they can adjust their transfer in both directions under uncertainty. Nevertheless, low types presumably have a higher tendency to respond with identity management and, likewise, high types are more prone to engage in self-deception. However, in our data we do not detect significant differences in the average scenario 1 back transfer across subjects who increase/decrease back transfers under uncertainty. It seems that identity management is not limited to low pro-sociality types and low as well as high types engage in self-deception.

Finally, we discuss the robustness of our results in terms of the design choices we made. It is known that the use of the strategy method may encourage reciprocal behavior due to experimenter demand effects (Zizzo, 2010). In fact, we find more conditional cooperation (reciprocators) among our subjects than in Fischbacher et al. (2001), yet still within the range of results in similar studies. However, there is no indication that our within subjects variation of the scenario biases behavior in any particular way. Moreover, it is worth to note that two of

---

[11]When excluding humpback-shaped and selfish reciprocators (mean of aggregate back transfers in scenario 1 less than 20), 60 out of 99 (61%) resisted.

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 engage in self-deception 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.

4   Conclusion

We conducted a modified trust game in order to analyze how reciprocators respond to systematic changes of self-image relevant factors. 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.

This behavioral pattern is particularly interesting for the trustees' back transfer choices. Uncertainty about implementation of the back transfer may not only be perceived as a situational excuse to behave more selfishly (self-deception) but may also be interpreted as an opportunity to invest in a pro-social self-image (identity management). The two predicted effects go in opposing directions. Our results show that twice as many subjects decrease than increase their transfers under uncertainty. It seems that self-deception is prevalent when subjects make the back transfer choice. However, some trustees do increase their back transfers with more uncertainty about the implementation. It appears that self-image concerns have an ambiguous nature, in the sense that self-signaling processes can go either way: via self-deception they can lead to less giving, via identity management they can induce more giving.

Are there characteristics that distinguish individuals who are prone to self-deception from those who may invest in identity? It seems reasonable to assume that individuals who give more (pro-social types) are more likely to engage in self-deception, while those who give less (pro-self types) tend to be generous in order to boost their ego. However, our analysis does not provide support for this. Self-deceiving behavior and identity management are both used across the entire spectrum of scenario 1 back transfers.

Finally, our evidence also shows that the effect of situational excuses extends beyond the setting of a dictator game where it has 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. See also Malmendier, te Velde and Weber (2014) or Regner (in press) for similar findings, while van der Weele et al. (2014) find no effect of a moral wiggle room manipulation in the context of reciprocity.

References

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and identity." *Quarterly Journal of Economics*, 115(3): 715–753.

Andreoni, James, and B Douglas Bernheim. 2009. "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects." *Econometrica*, 77(5): 1607–1636.

Aronson, Elliot. 1992. "The return of the repressed: Dissonance theory makes a comeback." *Psychological inquiry*, 3(4): 303–311.

Beauvois, JL, and RV Joule. 1996. "A radical theory of dissonance. European monographs in social psychology."

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and prosocial behavior." *American Economic Review*, 96(5): 1652–1678.

Bénabou, Roland, and Jean Tirole. 2011. "Identity, morals, and taboos: Beliefs as assets." *The Quarterly Journal of Economics*, 126(2): 805–855.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, reciprocity, and social history." *Games and Economic Behavior*, 10(1): 122–142.

Bodner, Ronit, and Drazen Prelec. 2003. "Self-signaling and diagnostic utility in everyday decision making." *The psychology of economic decisions*, 1: 105–26.

Dana, Jason, Roberto A Weber, and Jason Xi Kuang. 2007. "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness." *Economic Theory*, 33(1): 67–80.

Dufwenberg, M., and G. Kirchsteiger. 2004. "A theory of sequential reciprocity." *Games and Economic Behavior*, 47(2): 268–298.

Exley, Christine L. 2015. "Excusing selfishness in charitable giving: The role of risk." *The Review of Economic Studies*, 82(2): 587–628.

Falk, Armin, and Urs Fischbacher. 2006. "A theory of reciprocity." *Games and Economic Behavior*, 54(2): 293–315.

Feiler, Lauren. 2014. "Testing models of information avoidance with binary choice dictator games." *Journal of Economic Psychology*, 45: 253–267.

Festinger, Leon. 1962. *A theory of cognitive dissonance.* Vol. 2, Stanford university press.

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics*, 10(2): 171–178.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are people conditionally cooperative? Evidence from a public goods experiment." *Economics letters*, 71(3): 397–404.

Greiner, Ben. 2004. "The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics." Department of Economics, University of Cologne mimeo.

Grossman, Zachary. 2014. "Strategic ignorance and the robustness of social preferences." *Management Science*, 60(11): 2659–2665.

Grossman, Zachary. 2015. "Self-signaling and social-signaling in giving." *Journal of Economic Behavior & Organization*, 117: 26–39.

Grossman, Zachary, and Joel J van der Weele. 2017. "Self-image and willful ignorance in social decisions." *Journal of the European Economic Association*, 15(1): 173–217.

Haisley, Emily C, and Roberto A Weber. 2010. "Self-serving interpretations of ambiguity in other-regarding behavior." *Games and Economic Behavior*, 68(2): 614–625.

Hamman, John R, George Loewenstein, and Roberto A Weber. 2010. "Self-interest through delegation: An additional rationale for the principal-agent relationship." *The American Economic Review*, 100(4): 1826–1846.

Konow, James. 2000. "Fair shares: Accountability and cognitive dissonance in allocation decisions." *The American Economic Review*, 90(4): 1072–1091.

Larson, Tara, and C Monica Capra. 2009. "Exploiting moral wiggle room: Illusory preference for fairness? A comment." *Judgment and decision Making*, 4(6): 467–474.

Malmendier, Ulrike, Vera L te Velde, and Roberto A Weber. 2014. "Rethinking reciprocity." *Annu. Rev. Econ.*, 6(1): 849–874.

Matthey, Astrid, and Tobias Regner. 2011. "Do I really want to know? A cognitive dissonance-based explanation of other-regarding behavior." *Games*, 2(1): 114–135.

Mazar, Nina, On Amir, and Dan Ariely. 2008. "The dishonesty of honest people: A theory of self-concept maintenance." *Journal of Marketing Research*, 45(6): 633–644.

Murnighan, J Keith, John M Oesch, and Madan Pillutla. 2001. "Player types and self-impression management in dictatorship games: Two experiments." *Games and Economic Behavior*, 37(2): 388–414.

Regner, Tobias. in press. "Reciprocity under moral wiggle room: is it a preference or a constraint?" *Experimental Economics.*

Shalvi, Shaul, Francesca Gino, Rachel Barkan, and Shahar Ayal. 2015. "Self-serving justifications doing wrong and feeling moral." *Current Directions in Psychological Science*, 24(2): 125–130.

Spiekermann, Kai, and Arne Weiss. 2016. "Objective and subjective compliance: A norm-based explanation of moral wiggle room." *Games and Economic Behavior*, 96: 170–183.

Tirole, J, A Falk, and R Bénabou. 2016. "Narratives, Imperatives and Moral Reasoning." Mimeo.

Trautmann, Stefan T, and Ferdinand M Vieider. 2012. "Social influences on risk attitudes: Applications in economics." In *Handbook of risk theory.* 575–600. Springer.

van der Weele, Joël J. 2014. "Inconvenient truths: Determinants of strategic ignorance in moral dilemmas." *Available at SSRN 2247288.*

van der Weele, Joël J, and Ferdinand von Siemens. 2014. "Bracelets of pride and guilt? An experimental test of self-signaling in charitable giving."

van der Weele, Joël J, Julija Kulisa, Michael Kosfeld, and Guido Friebel. 2014. "Resisting moral wiggle room: how robust is reciprocal behavior?" *American economic Journal: microeconomics*, 6(3): 256–264.

Zizzo, Daniel John. 2010. "Experimenter demand effects in economic experiments." *Experimental Economics*, 13(1): 75–98.

Appendix

A. Screenshots of the decision interface
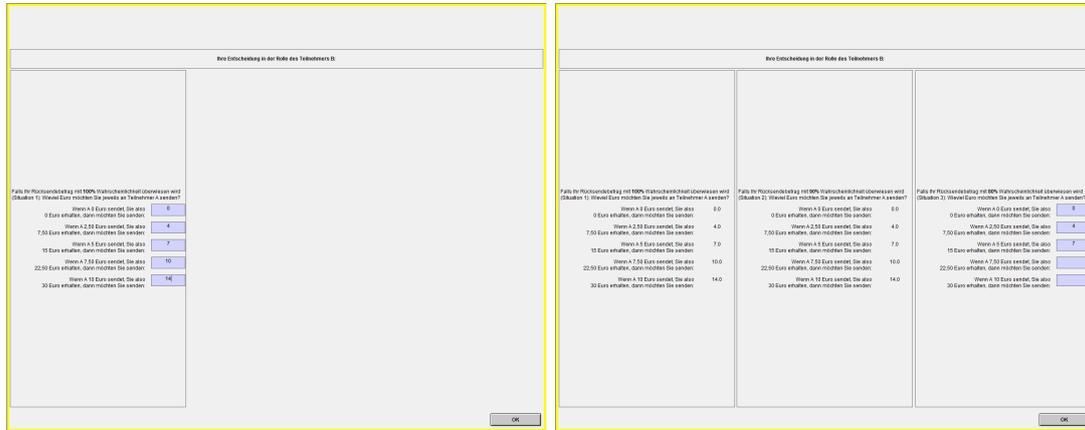
(a) Scenario 1   (b) Scenario 3

Figure 3: Screenshots of the decision interface (trustees enter their back transfer schedules subsequently for the scenarios with previous choices still visible)

B. Experimental instructions