

# JENA ECONOMIC RESEARCH PAPERS



# 2010 – 072

## Other-regarding behaviour: Testing guilt- and reciprocity-based models

**by**

**Tobias Regner**
**Nicole S. Harth**

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

© by the author.

Jena Economic Research Papers 2010 - 072

# Other-regarding behaviour: Testing guilt- and reciprocity-based models

Tobias Regner ♠* Nicole S. Harth ♣

♠ *Max Planck Institute of Economics, Strategic Interaction Group, Jena, Germany*

♣ *International Graduate College, Friedrich Schiller University, Jena, Germany*

January 31, 2011

**Abstract**

We analyse two types of belief-dependant models of social preferences: guilt aversion and reciprocity. In particular, we test the relevance of their input variables (second-order beliefs and general dispositions for guilt/reciprocity).

The data confirm the predictions of belief-dependant models. Both second-order beliefs and a participant's sensitivity to guilt/reciprocity are relevant for the decisions taken. Second-order beliefs appear to have an inverse U-shaped effect on the extent of kind behaviour.

*JEL classifications*: C91, D03, D84

*Keywords*: social preferences, other-regarding behaviour, belief-dependent models, experiments, trust game, guilt aversion, beliefs, psychological game theory, emotions, reciprocity

*Corresponding author (regner@econ.mpg.de)

We would like to thank seminar audiences at IMEBE 2010 in Bilbao, the 2010 ESA world meeting in Copenhagen and the 2010 IAREP/SABE conference in Cologne for their feedback. We are grateful to Giuseppe Attanasi, Jürgen Bracht, and Alexander Sebald for valuable comments. Alexander Raths provided excellent research assistance.

# 1 Introduction

Other-regarding behaviour is an established finding in the lab and the field. Yet, it is less explored what actually drives this behaviour. Outcome-based models of social preferences like Fehr and Schmidt (1999) or Bolton and Ockenfels (2000) explain transfers greater than zero with an aversion to inequity. Alternatively, belief-dependant models of social preferences use the psychological games framework of Geanakoplos, Pearce and Stacchetti (1989) to explain what makes people transfer more than they have to. This approach allows to consider various emotions or reciprocity. Models include Rabin (1993), Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006), Segal and Sobel (2007), and Battigalli and Dufwenberg (2007). They all have in common that whether I transfer more than necessary to another person or not, depends on my expectation about that person's expectation about my behaviour. Thus, predictions of such belief-dependant models are based on higher order beliefs (and their accurate, reliable elicitation) and a weighting parameter of the belief component that expresses how much emotions/intentions matter to the individual in the analysed context.

The paper focuses on two types of belief-dependant models (guilt aversion and reciprocity) and its aim is to test the relevance of their input variables: beliefs and the sensitivity of a person to experience, in our case, guilt or reciprocity. Besides dealing with different motivations the two models are distinct in the way second-order beliefs are related to behaviour. Take a trust game with sender $A$ and recipient $B$. The guilt aversion model (see for instance Charness and Dufwenberg, 2006) focuses on what ($B$ believes $A$ thinks) $A$ receives; second-order beliefs are a vehicle to express how much $B$ would be affected, if $B$ caused disappointment to $A$. Reciprocity (see for instance Dufwenberg and Kirchsteiger, 2004) is modeled considering what ($B$ believes $A$ thinks) $B$ receives; here second-order beliefs are a vehicle to express $B$'s dislike if $A$'s kindness were motivated by getting a high return, i.e., $B$ getting a low return. That

means increasing second-order beliefs have a positive effect on the amount $B$ returns in the guilt aversion[1] model, while they have a negative effect in the reciprocity[2] model.

In our experiment we use two games, a trust game and a mini trust game, to illustrate how these seemingly contradicting approaches work in parallel. In addition to incentivised elicitation of first- and second-order action beliefs, we assess participants' general dispositions (their sensitivity to feel guilt, and their attitude towards acting reciprocal) in the post-experimental questionnaire. This allows us to consider all model components (beliefs and sensitivities to guilt/reciprocity) and see how relevant they are for the decisions of participants. We also propose a novel way how to interpret perceived kindness as we implement a heterogeneous reference point based on trustees' first-order beliefs.

Our results are in line with previous findings of a correlation between beliefs and behaviour.[3] They largely confirm belief-dependant models: decisions of trustees are driven by i) expectations, ii) general dispositions and iii) expectations about expectations. Trustees tend to return less, the more their expectations about the transfer of the trustor are disappointed by the actual transfer; especially if they have a high general disposition to reciprocate negatively. Trustees tend to return more, the more their expectations are exceeded only in combination with a high attitude towards positive reciprocity. Trustees with a high sensitivity to feel guilt tend to cooperate more (in Game 1 where no feedback is given and thus expectations cannot be disappointed/exceeded). Last but not least, second-order beliefs appear to have an inverse U-shaped effect on the amount returned. Once they are higher than a certain level – our results suggest half of the available amount – the effect of increasing second-order beliefs on the amount returned changes from positive (as guilt aversion

[1] Guilt aversion's basic rationale is the following. The more I believe you were disappointed, the more guilt I would anticipate to feel. Hence, the more likely I am to take the kind choice to avoid the negative feeling that would result from the unkind choice.

[2] The basic reciprocity mechanism in Dufwenberg and Kirchsteiger (2004) could be described as follows. Generally, I am kind to you, if you are kind to me. But the more I believe you expect me to forgo a gain, the less I am willing to actually do it.

[3] See, among others, Dufwenberg and Gneezy (2000), Charness and Dufwenberg (2006), or Bacharach et al. (2007).

predicts) to negative (as reciprocity models predict).

The paper is organized as follows. Section 2 reviews the related literature. In section 3 we describe the experimental design and develop research hypotheses. Results are presented in section 4 and section 5 concludes.

## 2 Related Literature

People are considered to be reciprocal if they reward kind actions and punish unkind ones.[4] In belief-dependant models of reciprocity agents may derive utility from rewarding/punishing kind/unkind actions, even if this comes at a material cost. A key element is therefore how to assess whether an action has been kind/unkind, ideally from the perspective of the individual. This perceived kindness should then depend on i) the mere intentionality of an action and ii) the choice in the context of its alternatives. Both aspects have been tested empirically. Results in McCabe et al. (2003) and Falk et al. (2008) for instance confirm that it matters for recipients whether an action can really be attributed to the sender (in contrast to a random choice). Likewise, procedural concerns do play a role as shown by Bolton et al. (2005), for instance.

The focus of our paper is not on the questions of attribution or procedures. We take the general relevance of belief-dependant models for granted – based on these earlier findings – and focus our attention on the model parameters. We i) propose a new approach how to determine what is perceived as kindness, namely implementing a heterogeneous reference point based on first-order beliefs and ii) test the significance of the perceived kindness and the sensitivity to reciprocity in determining the returned amount. This sensitivity to reciprocity weighs the reciprocity term and affects whether the psychological benefit of being kind is large enough, i.e. whether one foregoes a higher material payoff. It is known that there are stable individual differences in people's attitude towards

[4] In more detail and to distinguish from other definitions we mean strong reciprocity, that is non-strategic behaviour unconditioned on future prospects. A reciprocal altruist (Trivers, 1971) would only reciprocate if there are future rewards arising from reciprocal actions.

acting reciprocal (Dohmen et al., 2009). Hence, an individual's sensitivity to reciprocity may be a relevant factor to explain differences in behaviour. Moreover, we do not treat positive and negative reciprocation as a general norm; instead we distinguish between positive and negative reciprocity following the psychological literature (Eisenberger et al., 2004).

The trust game serves as the workhorse experimental game for our analysis. Berg et al. (1995) conducted it first and it has been repeated numerous times. Fehr and Gächter (1998) survey trust games in the literature and conclude that never below 40% and sometimes above 60% of subjects exhibit reciprocal choices, while the fraction of subjects who behaves completely selfish lies between 20% and 30%. Costa-Gomes et al. (2010) provide evidence that the frequently found correlation between stated expectations and the level of trusting behaviour[5] is indeed to a large extent of causal nature, and hence in line with belief-dependant models.

In psychology "the prototypical cause of guilt would be the infliction of harm, loss, or distress on a relationship partner" (Baumeister et al., 1994). Elster (1998) introduced emotions to a broader audience among economists and guilt has probably received most attention. Ruffle (1999) and Dufwenberg (2002) applied the psychological games framework of Geanakoplos, Pearce and Stacchetti (1989) in order to integrate feelings of guilt into economic thinking. Charness and Dufwenberg (2006) use pre-play communication in a one-shot principal agent game to create a situation where guilt may arise, in particular when people make promises.[6] They find a positive effect of promises on beliefs and actions, and a positive correlation between second-order action beliefs and decisions. Battigalli and Dufwenberg (2007) provide a complementary theoretical model of guilt.

Several studies followed up on Charness and Dufwenberg (2006) to analyse

[5] See among others Dufwenberg and Gneezy (2000) or Bacharach et al. (2007).

[6] Other studies analysing the effect of guilt on behaviour include Ketelaar and Au (2003), Miettinen and Suetens (2008), Güth et al. (2009).

whether communication prior to play affects cooperative behavior in the way guilt aversion implies. Alternatively, acknowledged by Charness and Dufwenberg (2006), the correlation between beliefs and actions may not be caused by guilt feelings but rather by a false consensus effect (Ross et al., 1977). If causality is in fact reversed, not guilt aversion but for instance an aversion to lie (see Ellingsen and Johannesson, 2004) could explain the effect of promises. The results of Vanberg (2008) and Ellingsen et al. (2009) hint at the relevance of such alternative explanations. However, Bellemare et al. (2009) control for false consensus effects (which turn out to be substantial), and find that guilt aversion is still significant. Other studies that test the false consensus effect but still find evidence that beliefs cause behaviour include Fischbacher et al. (2001), Croson and Miller (2004), Frey and Meier (2004) and Reuben et al. (2009).

In this paper we do not try to analyse the direction of causality between beliefs and actions. We assume that beliefs cause behaviour based on these recent findings. Instead, we want to focus on the second element of a guilt aversion model, namely the sensitivity to feel guilt, that – to the best of our knowledge – has not received particular attention in analysis. According to Tangney (1995) individuals differ in the degree to which they are prone to feel shame and guilt. Hence, an individual's sensitivity to guilt may also explain differences in behaviour.

# 3 Method

## 3.1 Participants and Procedures

The experiment took place at the laboratory of the Max Planck Institute of Economics in Jena, Germany. 254 participants were recruited among students from various disciplines at the University of Jena 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 av-

erage, 75 minutes. The average earnings in the experiment have been €13.56 (including a €2.50 show-up fee).

At their 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 read the instructions. There was enough time to privately ask for clarifications about the instructions. Subjects had to pass several control questions before the experiment started, in order to make sure that they understood the instructions properly.

At the end of the experiment subjects were paid in cash according to their performance. Privacy was warranted during the payment phase.

## 3.2 Experimental Design

The experiment is composed of two different games: 1) a mini trust game and 2) a trust game. The sole interest of this paper are the decisions of the trustee, labeled participant $B$. The decisions of the trustor - participant $A$ - are of no particular interest for our analysis of belief-dependant models. The behaviour of $A$s is the topic of a companion paper, Harth and Regner (2009), that is written for a social psychology audience. In total 8 rounds were played. Participants knew that they were either participant $A$ or $B$. They were informed that Game 1 will be played in the first round and that from a later round onwards Game 2 may be played. Participants knew that they will play with a randomly selected other participant in each round.

### 3.2.1 Game 1

Game 1 is a mini trust game. Participant $A$ first chooses between an outside option (payoffs for $A$ and $B$: 6 experimental currency units (ECU), 4 ECU) and the investment. Participant $B$ was asked to choose between defection (payoffs: 0, 14) and cooperation (payoffs: 10, 10) independently of whether $A$ actually decided to invest. Neutral terms were used to label the decisions. Participants

knew that Game 1 or Game 2, the trust game, was going to be played for 8 periods in total. No feedback about period 1 choices was given. See Figure 1 for the structure of Game 1. In case participant $A$ cooperated in period 1, we slightly increased the outside option to a payoff of (8, 6) in the subsequent rounds of Game 1. Participants knew that the payoffs could be modified slightly after round 1.

[Figure 1 about here]

#### 3.2.2 Game 2

The second phase of the experiment consists of a trust game. When in Game 1 a participant $A$ decided to cooperate after a non-cooperative choice in a previous round, she started to play Game 2 from the next round onwards. In this standard trust game both participants ($A$ and $B$) had an endowment of 10 experimental currency units (ECU). The sender (participant $A$) had to decide how much to transfer to participant $B$. This amount ($x$) was doubled and added to $B$'s endowment. Then, $B$ decided how much of the available amount (10 ECU plus 2 times $x$) to return to $A$.

#### 3.2.3 Participants A

Those participants $A$ who did not cooperate in the first rounds of Game 1 were randomly allocated to one of the conditions of a 2 (guilt manipulation: yes vs. no) by 2 (feedback during Game 2: yes vs. no) - between-subjects-design. In the guilt manipulation condition we confronted participants $A$ who showed non-cooperative behavior in Game 1 with a message that appeared on the computer screen and was meant to induce guilt feelings. In the 'no feedback' condition participants $A$ were not informed about the Game 2 back transfer of participants $B$. Likewise, in the 'feedback' condition $A$s were informed about $B$s' back transfer at the end of the round. The guilt manipulation of participants $A$ or even its possibility was not announced before. Therefore, we exclude that this treatment variation can have any effect on $B$s. The other factor (participants

$A$ receive feedback during Game 2 or not) was known to participants and a potential effect on participants $B$ cannot be ruled out. However, testing the variables of interest does not show any significant difference between the two feedback conditions and we conclude that the $B$ decisions are not affected by the treatment variations implemented for the companion paper.

### 3.3 Beliefs

In order to elicit action beliefs we ask each participant $B$ in each period about her first-order and second-order belief. In game 1 this is the percentage of participants $A$ who they believe on average chooses RIGHT (i.e., first-order belief $\tau_B^A$), and the percentage - in her view - of participants $A$ who on average expect participant $B$ to choose RIGHT (i.e., $\tau_B^{AB}$).

In game 2 their first-order belief is the average transfer of participants $A$. During game 2 a second-order belief must be contingent on the actual transfer of participant $A$. This information is provided to participants $B$ right before the elicitation. They are also told that given this transfer participant $A$ knows what $B$ now has (endowment plus multiplied transfer). Then, they are asked what they expect $A$ expects to receive back from $B$.

Beliefs are collected as vectors of probabilities for the alternative choices with $\tau_k$ measuring the average belief of a player $k$. The "correctness" of the first-order beliefs will emerge from the comparison between beliefs and actual actions of participants $A$ in the respective period and session. Concerning the second-order beliefs their accuracy results from the comparison between second-order beliefs of $B$ and first-order beliefs of participant $A$ (e.g., $\tau_B^{AB}$ vs. $\tau_A^B$). Beliefs of $B$ were elicited in an incentive compatible fashion using a quadratic scoring rule (for an example, see Schotter and Sopher, 2007).[7] Great care was taken to make sure that participants understood the procedure.

[7] Belief elicitation requires quite some additional instructions, especially when incentivising belief statements and even more so when allowing beliefs to be probabilistic (see Artinger et al. (2010) for a survey). The fact that we experimentally enforce belief statements of course does not mean that participants naturally form such beliefs and are guided by them.

First-order beliefs of $A$ were elicited in the post-experimental questionnaire. They were not incentivised and only served the purpose of having a comparison for the second-order beliefs of $B$. Results from earlier sessions and a pilot were used for these payoff-determining comparisons.

## 3.4 Research Hypotheses

As illustrated in Battigalli and Dufwenberg (2009) the psychological games framework allows for the analysis of decisions from both the perspective of guilt aversion and from reciprocity. In the following we derive hypotheses for both models. Section 3.4.1 outlines what the benchmark guilt aversion model predicts for Game 1 and section 3.4.2 analyses Game 2 from the perspective of a reciprocity model. In section 3.4.3 we discuss what to expect when both models are considered in parallel.

### 3.4.1 Guilt Aversion

Similar to Charness and Dufwenberg (2006) guilt aversion would predict the following for Game 1. Let $\tau_A^B$ be $A$'s belief about the probability that $B$ picks RIGHT. Then $\tau_B^{AB}$ denotes $B$'s belief regarding $\tau_A^B$. In order to measure the amount $B$ thinks she hurts $A$ by picking LEFT, we calculate the difference between $A$'s payoff when $B$ plays RIGHT and when $B$ plays LEFT (weighted by the second-order belief $\tau_B^{AB}$): $10 \cdot \tau_B^{AB} - 0 = 10 \cdot \tau_B^{AB}$

How much this actually affects $B$ is expressed by taking her sensitivity to guilt $\gamma_B$ into account. Hence, if $B$ selects LEFT, she therefore experiences guilt of $10 \cdot \tau_B^{AB} \cdot \gamma_B$. This psychological cost of guilt reduces $B$'s material payoff of choosing LEFT. Given $B$ is rational she will prefer RIGHT over LEFT if the following inequity holds:

$$U_B^{LEFT} = 14 - 10 \cdot \tau_B^{AB} \cdot \gamma_B \quad < \quad 10 = U_B^{RIGHT} \tag{1}$$

Note that $\gamma_B = 0$ represents the model's special case of pure self-interest. We derive the following two hypotheses for behaviour in Game 1:

**Hypothesis 1** *(Game 1) The higher B's second-order belief $\tau_B^{AB}$ is, the higher is the probability that B will choose RIGHT.*

**Hypothesis 2** *(Game 1) The higher B's sensitivity to guilt $\gamma_B$ is, the higher is the probability that B will choose RIGHT.*

### 3.4.2 Reciprocity

We now turn to Game 2. Generally, people are considered to be reciprocal if they reward kind actions and punish unkind ones. The creation of utility by matching the signs of kindness and perceived kindness may be regarded as a key element of the sequential reciprocity model in Dufwenberg and Kirchsteiger (2004). Reciprocation means responding to positive perceived kindness of someone with positive kindness of oneself, and to negative perceived kindness with negative kindness.

Hence, utility is expressed by the material payoff $\pi$ and the additional reciprocity term consisting of the sensitivity to reciprocity $\alpha$, kindness $\kappa$ and perceived kindness $\lambda$ (simplified notation): $U = \pi + \alpha \cdot \kappa_{AB} \cdot \lambda_{BAB}$

Kindness is one's reply to perceived kindness (how kind one perceives someone else to be). This perceived kindness depends on second-order beliefs. It is considered how $B$'s own payoff changes depending on the second-order belief, that is the higher the second-order belief the less $B$ gets and thus the lower is the perceived kindness of $A$. The reciprocity term is then weighted by one's sensitivity to reciprocity and when it is large enough it outweighs the material loss compared to a less kind option.

In order to measure (perceived) kindness Dufwenberg and Kirchsteiger (2004) use a reference point: the equitable payoffs which is usually the average of the available payoffs. Kindness is then defined as the discrepancy between the payoffs resulting from the actual choice and the equitable payoff. The perceived

kindness depends on the payoffs in expectations, that is the available choices are weighted by their assumed likelihood (the second-order beliefs). Perceived kindness is therefore defined as the discrepancy between these probability-weighted payoffs and the equitable payoff.

It is absolutely plausible to assume that the average represents a general reference point to distinguish kind and unkind actions. Yet – in the spirit of Manski (2004) – if known we may use what individuals actually believe[8] as a reference for kindness, that is their expected action (first-order belief) at the beginning of the game. Following the psychological literature (Zeelenberg et al., 2000) anything that is beyond one's expectations will be seen as kind (a positive surprise), and anything below of what is expected will be regarded as unkind (a disappointment).

Utility of $B$ is then expressed by the material payoff $\pi_B$ and the additional reciprocity term. The sensitivity to reciprocity $\alpha_B$ is assumed to be exogenous but individually heterogenous. Kindness $\kappa_{AB}$ is assumed to sign-match perceived kindness $\lambda_{BAB}$, which is determined by $\delta_{AB}$ (the action of $A$ in reference to $B$'s initial expectation, expressing whether $A$ is perceived as kind or not) and $B$'s second-order belief $\tau_B^{AB}$ ($B$'s thoughts about why $A$ may have been kind).

$$U_B \quad = \quad \pi_B + \alpha_B \cdot \kappa_{AB} \cdot \lambda_{BAB}(\delta_{AB}, \tau_B^{AB}) \tag{2}$$

Note that for $\alpha_B = 0$ the reciprocity term disappears and we get the special case of pure self-interest. We derive the following hypotheses from equation 2 for behaviour in Game 2:

**Hypothesis 3** *(Game 2) The higher $B$'s second-order belief $\tau_B^{AB}$ is, the lower is the amount $B$ sends back.*

Following evidence from social psychology (see, for instance, Eisenberger et

[8] Also in the context of social preferences, Bellemare et al. (2008) take a similar approach and demonstrate that incorporating subjective probabilities improves predictions of the inequity aversion model of Fehr and Schmidt (1999).

al., 2004) we distinguish between the sensitivity to positive ($\alpha_B$) and negative ($\beta_B$) reciprocity, slightly modifying the model. For the case of $\delta_{AB} > 0$ or a positive surprise, only the sensitivity to positive reciprocity should matter.

**Hypothesis 4** *(Game 2) The higher B's sensitivity to positive reciprocity $\alpha_B$ is, the higher is the amount B sends back.*

For the case of $\delta_{AB} < 0$ or a disappointment, only the sensitivity to negative reciprocity should matter. In equation 2 $\beta_B$ replaces $\alpha_B$.

**Hypothesis 5** *(Game 2) The higher B's sensitivity to negative reciprocity $\beta_B$ is, the lower is the amount B sends back.*

Finally, how much $B$ got positively surprised/disappointed by the action of $A$ should have a moderating effect on someone's attitude to act reciprocal.

**Hypothesis 6** *(Game 2) The discrepancy $\delta_{AB}$ between B's first-order belief and the actual transfer received from A moderates the effect of one's sensitivity to reciprocity.*

#### 3.4.3 Combined model of guilt aversion and reciprocity

We applied the guilt aversion model to Game 1 and the reciprocity model to Game 2. Yet, both motivations may actually play a role in either game. It should be interesting to analyse the particular decision situations taking into account that second-order beliefs have opposing effects (positive for guilt aversion, negative for reciprocity) on the amount to be sent.

Applying reciprocity to Game 1 predicts a negative effect of second-order beliefs on cooperative behaviour. $B$'s kindness of playing RIGHT is $\kappa_{AB} = 10 - 5 = 5$ (using an equitable payoff of 5). The perceived kindness of $A$ depends on second-order beliefs given $A$ played right, and what $B$ gets when $A$ plays left:

$\lambda_{BAB}(\tau_B^{AB}) = (1-\tau) \cdot 14 + \tau \cdot 10 - (((1-\tau) \cdot 14 + \tau \cdot 10) + 4) \cdot 1/2 = 5 - 2\tau$

In the context of reciprocity $B$'s own payoffs are weighted by second-order beliefs (14 if $B$ plays LEFT, 10 if $B$ chooses RIGHT). In contrast, guilt aversion considers $A$'s payoffs (0 if $B$ plays LEFT, 10 if $B$ chooses RIGHT). Thus, the reciprocity model focuses on what ($B$ believes $A$ thinks) $B$ receives, while the guilt aversion model focuses on what ($B$ believes $A$ thinks) $A$ receives. It is worth to note that in Game 1 $B$ has less to lose when he picks RIGHT instead of LEFT (10-14), than $A$ would gain (10-0). Hence, one may expect the effect of guilt aversion (via the believed change in $A$'s payoffs) to be dominant. Also, the design of Game 1 limits $B$ to either an opportunistic (LEFT) or a fair (RIGHT) choice. $B$ cannot give more to $A$ than he keeps to himself, and he cannot consider such an outcome in his second-order beliefs.

Game 2, on the other hand, does not restrict the choice set of $B$: whatever is available after $A$'s transfer can be returned. This means $B$ may believe that $A$ expects to get back more than $B$ would actually keep himself. Thus, Game 2 allows second-order beliefs in a domain that was out of bounds in Game 1. Also, in contrast to Game 1, the design of Game 2 is symmetric in the payoffs. Every ECU that $B$ adds to the amount he returns will end up in $A$'s pocket. This leads to the intriguing question which effect of second-order beliefs on the amount returned (via guilt aversion's change in $A$'s payoffs or reciprocity's change in $B$'s payoffs) prevails. Is the negative effect of second-order beliefs proposed by the reciprocity model dominant? Or is the positive effect of second-order beliefs suggested by the guilt aversion model substantial enough to cancel or possibly outweigh the reciprocity effect? Another possibility is that both effects matter in different ranges of second-order beliefs. This would be in line with Attanasi and Nagel (2009) who find evidence for a positive relationship between second-order beliefs and transfers to another participant for low and medium ranges of second-order beliefs, and a negative relationship for high levels of second-order beliefs. This may point to a dominance of the guilt aversion effect for belief ranges where $A$ would end up with (close to) nothing. Possibly only

when second-order beliefs are in a relatively high range $B$'s focus shifts on the resulting loss to himself. Hypothesis 7 captures this alternative approach.

**Hypothesis 7** *$B$'s second-order beliefs $\tau_B^{AB}$ have a positive effect on the amount $B$ sends back for low levels of $\tau_B^{AB}$, while they have a negative effect on the amount $B$ sends back for high levels of $\tau_B^{AB}$.*

We will attempt to model such a relationship with an additional squared term of second-order beliefs.

# 4 Results

Participants first played Game 1 (mini trust game) and then Game 2 (trust game). Everybody played Game 1 during the first two periods, then participation gradually decreased (see Figure 2 for details). Switching from Game 1 to Game 2 was determined by the actions of participants $A$. We employed a random-stranger matching procedure, thus it was by chance whether participants $B$ moved into Game 2.

[Figure 2 about here]

The post-experimental questionnaire contained questions about participants' sensitivity to feel guilt, and their attitude towards acting reciprocal on a scale from 1 to 7. Several tests have been developed by psychologists to measure guilt, and most correlate highly (Kugler and Jones, 1992). For this study we chose a very short one that assesses the ease with which guilt is generally experienced (Moulton et al., 1966). The two/two questions about positive/negative reciprocity were aggregated to one/one value (Cronbach's $\alpha = 0.64/0.67$) and Figure 3 shows the histograms for the sensitivity to experience *Guilt*, positive (*PosRec*), and negative reciprocity (*NegRec*). The distributions for reciprocity are fairly similar to the ones in Dohmen et al. (2009) who aggregated three questions each and used data of the 2005 wave of the SOEP, a large representative survey of German households. It is noteworthy that, as in Dohmen et al.

(2009), we do not find that sensitivity to positive and negative reciprocity are strongly correlated. They seem to be different constructs (see Eisenberger, et al. 2004). In addition, we asked participants how relevant the opinion of others is to them (*OtherOpinion*), and how important it is for them to have and comply with certain principles in life (*Principles*). We also asked for some background information (age, gender).

[Figure 3 about here]

## 4.1 Game 1

In Game 1 the choice of participants $B$ consisted of selecting whether to cooperate (resulting in a payoff of 10 for both) or not ($A$ receives nothing, $B$ gets 14). $B$s knew that i) they had to choose independently of whether $A$ actually decided to invest or not, but ii) their choice only mattered when $A$ chose right. As described before this setup creates a situation where individuals who cooperate may do so because they want to avoid feelings of guilt. No feedback was provided after a period. The cooperation rate of participants $B$ over the course of Game 1 is fairly stable, see Figure 4.

[Figure 4 about here]

Table 1 shows the results of two random effects Probit regressions. The dependant variable is whether participant $B$ cooperated (1) or not (0). Column I shows results for an unbalanced panel model based on periods 1 to 7 ($N = 587$). The sort out of participants $B$ during the course of Game 1 is entirely based on behaviour of participants $A$. Hence, there is no indication that sample selection effects after period 3 would play a role.

The coefficients for the second-order belief as well as for the sensitivity to feel guilt are positive and highly significant. Hypotheses 1 and 2 cannot be rejected. None of the control variables is significant at the 5%-level.

Column II contains results for a combined model of guilt aversion and reciprocity. Thus, we add a squared term of second-order beliefs and the sensi-

Table 1: Impact on the Cooperation Rate (Game 1)

| DV: Game 1 cooperation | I: guilt aversion | | II: combined model | |
|---|---|---|---|---|
| | coeff. | st.error | coeff. | st.error |
| 2nd order belief sq. | – | – | -.00024 | .00017 |
| 2nd order belief | .0141 | .0049 *** | .0378 | .0185 ** |
| Guilt | .3501 | .1349 *** | .2713 | .1311 ** |
| PosRec | – | – | .2829 | .2042 |
| NegRec | – | – | -.3068 | .1233 ** |
| Principles | .0748 | .1541 | .0126 | .1481 |
| OtherOpinion | .0645 | .1344 | .0612 | .1264 |
| female | -.5861 | .3258 * | -.4597 | .3071 |
| age | -.0338 | .0535 | -.0134 | .0512 |
| constant | -1.951 | 1.667 | -2.791 | 1.858 |
| $N$ | 587 | | 587 | |
| Log likelihood | -310.04 | | -305.41 | |

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

tivities towards positive/negative reciprocity. The squared term's coefficient is not significantly different from zero, while the linear term remains positive and significant. Hence, it does not appear that the reciprocity model's negative effect (via second-order beliefs) affects cooperation behaviour in Game 1. The coefficient for sensitivity of guilt is still significant as well as the one for attitude towards negative reciprocity.

## 4.2 Game 2

In Game 2 – a standard trust game – participants $B$ are i) asked about their first-order beliefs, ii) they are told what their randomly matched trustor sent them, iii) they are asked to tell us their second-order beliefs (based on what $A$ sent), and then iv) they decide what to return to $A$. Generally, the trust game allows trustees to reciprocate: The more one receives, the more one may be inclined to return. Following belief-dependant models the returned amount actually depends on the beliefs of recipients and our particular design enables us to take this into account and understand better, why they reciprocate.

There are 423 observations for Game 2. Figure 5 shows how much $A$s sent in Game 2. The amount sent and the amount returned are highly correlated, a

common finding in trust games.

[Figure 5 about here]

Again an unbalanced random effects model takes individual heterogeneity into account. We use a Tobit model since the amount returned is limited to the range between 0 and 30. Second-order beliefs are provided as a percentage of the actually available amount, since Bs were informed about what has been sent to them before they were asked for second-order beliefs. The difference $\delta$ between the amount sent from $A$ to $B$ and the first-order belief of $B$ is calculated to express, whether $B$ is positively surprised ($\delta > 0$) or disappointed ($\delta < 0$). 17 times the participant expected just what was sent to him/her and the difference was zero. Figure 6 shows the distribution of the difference between the amount sent from $A$ to $B$ and the first-order belief of $B$. Table 2 column I shows the regression results.

[Figure 6 about here]

The coefficient for second-order beliefs is positive and highly significant. Also the difference between actual amount sent and the expectation about it – expressing whether a participant has reason to be positively surprised (high values of $\delta$) or disappointed (low values) – seems to have a significant positive effect on the amount returned. Our measures for the sensitivity of positive/negative reciprocity do not seem to have an effect, nor any of the control variables. The lack of significance of the reciprocity measures is not very surprising, though. As can be seen in Figure 3 and pointed out before positive and negative reciprocity cannot be regarded as symmetric concepts, see also Dohmen et al. (2009).

Hence, it may be more appropriate to split the sample depending on whether $\delta < 0$ (the participant should be disappointed and negative reciprocity should matter) or $\delta > 0$ (the participant should be positively surprised and positive reciprocity should matter). Table 2 columns II and III contain the results for these split samples.

Table 2: Game 2 (reciprocity model)

| DV: amount returned | I: all obs | | II: $\delta < 0$ | | III: $\delta > 0$ | |
|---|---|---|---|---|---|---|
| | coeff. | st.error | coeff. | st.error | coeff. | st.error |
| 2nd order belief | 7.337 | 1.323 *** | 7.888 | 1.761 *** | 17.16 | 2.675 *** |
| Delta | .8988 | .4989 * | 2.281 | .7059 *** | -.7073 | .8644 |
| PosRec | 1.111 | .6553 | – | – | .6544 | .8644 |
| PosRec * Delta | .0734 | .0711 | – | – | .2106 | .1467 |
| NegRec | -.1039 | .4343 | -1.526 | .6007 ** | – | – |
| NegRec * Delta | -.0841 | .0526 | -.3596 | .1877 * | – | – |
| Principles | -.8448 | .5107 * | .2228 | .4196 | -.8076 | .5475 |
| OtherOpinion | .335 | .4513 | .5978 | .4011 | .5277 | .4853 |
| feedback | -.2238 | 1.011 | -1.214 | .8487 | -.236 | 1.112 |
| female | -.804 | 1.112 | .527 | .9623 | -.6596 | 1.233 |
| age | -.1643 | .1896 | .1195 | .1536 | -.2306 | .2138 |
| constant | 6.286 | 5.794 | -7.265 | 5.009 | 5.322 | 6.392 |
| $N$ | 423 | | 201 | | 205 | |
| Log likelihood | -789.23 | | -274.54 | | -483.18 | |

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

When participants should experience some kind of disappointment ($\delta < 0$, column II) second-order beliefs are still highly significant as well as the difference between actual amount sent and the expectation about it. High values of negative reciprocity seem to have a negative effect on the amount sent (statistically significant at the 5%-level). The significance of the interaction term between negative reciprocity and the difference is marginally significant. None of the control variables are significant.

When participants should be positively surprised ($\delta > 0$, column III) second-order beliefs are also highly significant. The coefficient of positive reciprocity is positive, but the effect is not statistically significant. Likewise, the difference between actual amount sent and the expectation about it and the interaction term between positive reciprocity and the difference do not seem to have an effect. None of the control variables are significant.

Hypothesis 3 cannot be rejected on the basis of the estimates of the reciprocity model in table 2. The supposed negative effect of second-order beliefs is in fact positive and highly significant. This may indicate the relevance of motivation in accordance to the guilt aversion model. Table 3 presents results of the

combined model that features an additional squared term of second-order beliefs to account for the potential positive/negative effect of second-order beliefs on the amount returned in low/high ranges of second-order beliefs.

Table 3: Game 2 (combined model)

| DV: amount returned | I: all obs | | II: $\delta < 0$ | | III: $\delta > 0$ | |
|---|---|---|---|---|---|---|
| | coeff. | st.error | coeff. | st.error | coeff. | st.error |
| 2nd order belief sq. | -30.52 | 6.621 *** | -41.32 | 8.733 *** | -30.22 | 12.13 ** |
| 2nd order belief | 34.62 | 5.308 *** | 43.21 | 6.943 *** | 39.24 | 9.323 *** |
| Delta | .4201 | .5191 | 1.974 | .7087 *** | -1.263 | .8961 |
| PosRec | .6349 | .6061 | – | – | .0585 | .8774 |
| PosRec * Delta | .1117 | .0742 | – | – | .2998 | .1517 ** |
| NegRec | -.2163 | .3871 | -1.55 | .5313 *** | – | – |
| NegRec * Delta | -.0764 | .0539 | -.4021 | .1863 ** | – | – |
| Guilt | .4479 | .4405 | .2321 | .3717 | .3549 | .5482 |
| Principles | -.8648 | .4569 * | .0053 | .3756 | -.8184 | .5294 |
| OtherOpinion | .2413 | .4037 | .3493 | .3671 | .3355 | .4668 |
| feedback | .1679 | .8908 | .1548 | .7182 | -.1541 | 1.048 |
| female | -1.185 | .9993 | .079 | .815 | -.6373 | 1.17 |
| age | -.0762 | .1677 | .2161 | .1292 * | -.1319 | .203 |
| constant | -1.441 | 5.378 | -14.22 | 4.376 *** | -.9829 | 6.487 |
| $N$ | 423 | | 201 | | 205 | |
| Log likelihood | -764.86 | | -251.61 | | -479.50 | |

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

Generally, the combined model appears to fit the data better as likelihood ratio tests for all three specifications (columns I-III in table 2/3, respectively) prefer the combined model at least at the 5%-level. The coefficient of the squared term of second-order beliefs is negative and the one for the linear term is positive. Both are significant at least at the 2%-level in all specifications. As in the reciprocity model in table 2 neither $\delta$ nor the measure for attitude towards positive/negative reciprocity have a significant effect in the full sample (column I).

When participants should experience some kind of disappointment ($\delta < 0$, column II) the difference between actual amount sent and the expectation about it has a positive and highly significant effect. The measure for the sensitivity to negative reciprocity has a negative and highly significant coefficient. The interaction term between these two is negative and significant at the 5%-level.

When participants should be positively surprised ($\delta > 0$, column III) the interaction term between $\delta$ and the measure for attitude towards positive reciprocity is significantly positive at the 5%-level.

Overall, there is strong evidence for a negative effect of the attitude towards negative reciprocity on the amount returned. Hypothesis 5 cannot be rejected. On the other hand, we do not find evidence for an effect of the attitude towards positive reciprocity, and we have to reject hypothesis 5. However, in the positive as well as in the negative domain we find an interaction effect of the difference between actual amount sent and the expectation about it and the respective attitude towards positive/negative reciprocity. We cannot reject hypothesis 6. Last but not least, we find strong evidence in favour of hypothesis 7 as $B$'s second-order beliefs $\tau_B^{AB}$ have a positive effect on the amount $B$ sends back for low levels of $\tau_B^{AB}$, while they have a negative effect for high levels of $\tau_B^{AB}$.

All results are robust to specifications that use a panel OLS model.

### 4.3 Discussion

Analysis of Game 1 is in line with the predictions of guilt aversion. Both second-order beliefs and the sensitivity to feel guilt seem to explain the decisions of participants. A combined model of guilt aversion and reciprocity does not appear to be a better specification.

In Game 2, in contrast to the prediction of a reciprocity-based model, the coefficient of second-order beliefs is positive (and highly significant). The specification featuring an additional squared term of second-order beliefs appears to be a better fit. Such a combined model could be interpreted as putting more weight on guilt aversion as a motivation when the level of $B$'s second-order beliefs is rather low, while putting more weight on reciprocity when the level of $B$'s second-order beliefs is rather high.

When $B$ decides how much to return in Game 2 every ECU transferred signifies one ECU less for $B$. Hence, for low levels of second-order beliefs (and therefore a low/high payoff for $A/B$) it seems that the effect of an additional

ECU for $A$ dominates the effect of one ECU less for $B$. Only for high levels of second-order beliefs (and therefore a high/low payoff for $A/B$) it seems that the effect of an additional ECU for $A$ is dominated by the effect of one ECU less for $B$.

Figure 7 illustrates the inverse U-shaped effect of second-order beliefs based on the estimations of the combined model in table 3. According to the estimations the ceteris paribus effect of increasing second-order beliefs starts to become negative around 0.5-0.6. Of the amount $B$ has available after the transfer of $A$ she will return more the higher her second-order beliefs are, as long as she does not think $A$ expects her to return more than half. When she does, she will return less the higher her second-order beliefs are. Among our observations second-order beliefs of more than 0.5 are somewhat rare (around 10%). As mentioned before the general tendency of a decreasing effect of second-order beliefs if they go beyond a certain threshold has also been found by Attanasi and Nagel (2009) who use a within-subject design. Interestingly, a similar pattern is found by Fehr and Rockenbach (2003) in a standard trust game. The desired back transfer of trustors is communicated to trustees, and classified into low/high when the trustor would earn less/more than the trustee. In the condition that is comparable to our design (their no-fine-possible condition) the actual back transfer is lower for high desired back transfers, although it is not clear whether the difference is significant (the article focuses on the condition where a fine can be imposed). Second-order beliefs in Game 1 are capped at an equal split (10, 10). This may be the reason why no decreasing effect of second-order beliefs is found in Game 1.

[Figure 7 about here]

Besides its effect via second-order beliefs we also analyse reciprocity in the combined model by looking at the individual reference point for kindness, their first-order beliefs. The difference $\delta$ between actual amount sent from $A$ to $B$

and $B$'s expectation about it (first-order belief), is highly significant under disappointment ($\delta < 0$), but there is no significance when $B$ is positively surprised ($\delta > 0$). Similarly, we find a main effect of the sensitivity to negative reciprocity, but no significance of the sensitivity to positive reciprocity. However, the impact of $\delta$ is moderated by the sensitivity to positive/negative reciprocity no matter whether $\delta$ is greater or less than zero.

This asymmetry with respect to positively and negatively reciprocal behaviour adds to the list of findings of that kind (for instance Blount (1995), Gneezy et al. (2000), Offerman (2002), Kube et al. (2006), Falk et al. (2008), Dohmen et al. (2009), Al-Ubaydli and Lee (2009)). These studies show that people do reciprocate negatively, but they do much less often reciprocate positively, if at all. We also observe this type of behaviour and in addition, we connect it to a model input variable (the sensitivity to positive/negative reciprocity) that is individually heterogenous. This may explain the differences in behaviour. Belief-dependant models of reciprocity (Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006)) use a single parameter to express an individual's sensitivity to reciprocity, assuming it is a single trait. This appears to be too generalising based on the different distributions and lack of correlation between the sensitivity to positive/negative reciprocity, and the different effects they have.

# 5 Conclusions

Other-regarding behaviour is an established finding in the lab and the field. It is less clear what actually drives this behaviour. We test the predictions of two types of belief-dependant models of social preferences: guilt aversion and reciprocity. This strand of models explains other-regarding behaviour building on the psychological games framework of Geanakoplos, Pearce and Stacchetti (1989). In contrast to outcome-based models (for instance Fehr and Schmidt (1999) or Bolton and Ockenfels (2000)) this approach does not assume payoffs

to be relevant for decision making. It incorporates higher order beliefs and actions into the utility function to allow for the consideration of emotions and reciprocity.

An additional squared term of *second-order beliefs* combines the motivations of guilt aversion (positive effect of second-order beliefs on other-regarding behaviour) and reciprocity (negative effect). This specification confirms an inverse U-shaped effect of second-order beliefs on the amount returned. For low levels of second-order beliefs (and therefore a low/high payoff for $A/B$) it seems that the effect of an additional payoff unit for someone else dominates the effect of the own loss of that payoff unit. Only for high levels of second-order beliefs (and therefore a high/low payoff for $A/B$) it seems that the effect of an additional payoff unit for someone else is dominated by the effect of the own loss. Our model estimates suggest that increasing second-order beliefs have a ceteris paribus positive effect on the amount returned as long as one thinks the other expects one to return less than half. When one thinks the other expects one to return more than half, it seems increasing second-order beliefs start to have a ceteris paribus negative effect on the amount returned.

While evidence for reciprocity via a negative effect of second-order beliefs appears to be limited to high ranges of second-order beliefs, we do find strong evidence for reciprocity using as well *first-order beliefs* to express perceived kindness. These expectations provide a heterogenous reference point. Anything beyond them is seen as kind (a positive surprise), and anything below is regarded as unkind (a disappointment). Trustees tend to return less, the more their expectations about the transfer of the trustor are disappointed by the actual transfer. Also *general dispositions* play a substantial role as the effect of disappointed expectations is particularly strong with a high general disposition to reciprocate negatively. Trustees tend to return more, the more their expectations are exceeded, but only in combination with a high attitude towards positive reciprocity. Trustees with a high sensitivity to feel guilt tend to cooperate more (in Game 1 where no feedback is given and thus expectations

cannot be disappointed/exceeded).

Overall, our study confirms the belief-dependant approach to model social preferences. All important model components turn out to be significant.[9] The inverse U-shaped effect of second-order beliefs on the amount returned combines the motivations guilt aversion and reciprocity are supposed to have on behaviour. This is shown in Game 2, a continuous trust game, where trustees can distribute the entire available amount and payoffs between trustor and trustee are exchanged at an equal rate. In a situation with a limited choice set and unequal exchange rate one of the effects may globally dominate. This is what we find in Game 1. It should be interesting to see how further research on the relationship between beliefs and decision making can contribute to a better understanding of human behaviour.

[9] As mentioned before our design does not aim to control for false consensus effects. These may bias results, but studies of Bellemare et al. (2009) or Costa-Gomes et al. (2010), for instance, show that a causal relationship between beliefs and behaviour in trust games persists after controlling for false consensus effects.

## References

Al-Ubaydli, O. and Lee, M. S. (2009). An experimental study of asymmetric reciprocity. *Journal of Economic Behavior and Organization*, **72** (2), 738-749

Attanasi, G. and Nagel, R. (2009). Actions, Beliefs and Feelings: An Experimental Study on Dynamic Psychological Games. *working paper*

Artinger, F., Exadaktylos, F., Koppel, H. and L. Sääksvuori (2010). Applying Quadratic Scoring Rule in multiple choice settings. *Jena Economic Research Papers*, **Vol. 4**, 2010-021.

Bacharach, M., Guerra, G. and D.J. Zizzo (2007). The self-fulfilling property of trust: An experimental study. *Theory and Decision*, **Vol. 63**, 349-388.

Battigalli, P., and M. Dufwenberg M. (2007). Guilt in Games. *American Economic Review Papers and Proceedings*, **97**, 170-176

Battigalli, P., and M. Dufwenberg M. (2009). Dynamic Psychological Games. *Journal of Economic Theory*, **144**, 1-35

Baumeister, R., Stillwell, A., and Heatherton, T. (1994). Guilt: An Interpersonal Approach. *Psychological Bulletin*, **115**, 243–267.

Bellemare, C., Kröger, S., and van Soest, A. (2008). Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica*, **76** (4), 815–839.

Bellemare, C., Sebald, A., and Strobel, M. (2009). Measuring the Willingness to Pay to Avoid Guilt: Estimation using Equilibrium and Stated Belief Models. *forthcoming: Journal of Applied Econometrics*

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, Reciprocity and Social History. *Games and Economic Behavior*, **10**, 122–142.

Blount, S. (1995). When social outcomes arent fair: the effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes*, **63**, 131–144.

Bolton, G., Brandts, J., and A. Ockenfels (2000). Fair procedures: evidence from games involving lotteries. *Economic Journal*, **115**, 1054-1076

Bolton, G. and A. Ockenfels (2000). A theory of equity, reciprocity and competition. *American Economic Review*, **100**, 166-193

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, **74** (6), 1579–1601.

Costa-Gomes, M. A. , S. Huck and G. Weizsäcker (2010). Beliefs and Actions in the Trust Game: Creating Instrumental Variables to Estimate the Causal Effect. *IZA discussion paper*, No. 4709.

Croson, R. and and Miller, M. (2004). Explaining the Relationship between Actions and Beliefs: Projection vs. Reaction. *mimeo*

Dohmen, T. , A. Falk, D. Huffman and U. Sunde (2009). Homo Reciprocans: Survey Evidence on Behavioral Outcomes. *Economic Journal*, **119**, 592–612.

Dufwenberg, M. (2002). Marital investments, time consistency and emotions. *Journal of Economic Behavior and Organization*, **48**, 57–69.

Dufwenberg, M., and U. Gneezy (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior*. 30, 163-182.

Dufwenberg, M., and G. Kirchsteiger (2004). A Theory of Sequential Reciprocity. *Games and Economic Behavior*. 47, 268-298.

Eisenberger, R., Lynch, P., Aselage, J., and Rohdieck, S. (2004). Who Takes the most Revenge? Individual Differences in Negative Reciprocity Norm Endorsement. *Personality and Social Psychology Bulletin*, **30**, 787–799.

Ellingsen, T., and M. Johannesson (2004). Promises, Threats and Fairness. *Economic Journal*, **114**, 397–420.

Ellingsen, T., Johannesson, M., Tjotta, S., and Torsvik, G. (2009). Testing guilt aversion. *Games and Economic Behavior*, **68**(1), 95–107.

Falk, A., Fehr, E., and U. Fischbacher (2008) Testing theories of fairness intentions matter. *Games and Economic Behavior*, **62**(2), 287–303.

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

Fehr, E. and B. Rockenbach (2003). Detrimental effects of sanctions on human altruism. *Nature*, **422**, 137-140

Fehr, E. and K. Schmidt (1999). A Theory of Fairness, Competition and Cooperation. *Quarterly Journal of Economics*, **114**, 817-868

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

Frey, B. and S. Meier (2004). Social comparisons and pro-social behavior: Testing 'conditional cooperation' in a field experiment. *Quarterly Journal of Economics*, **114**, 817-868 *American Economic Review*, **94**, 1717–1722.

Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, **1**, 60–79.

Greiner, B. (2004). An Online Recruitment System for Economic Experiments. in: K. Kremer and V. Macho, editors, *Forschung und wissenschaftliches Rechnen 2003*. GWDG Bericht 63, Göttingen : Ges. für Wiss. Datenverarbeitung.

Güth, W., Ploner, M., and Regner, T. (2009). Determinants of in-group-bias: Is group affiliation mediated by guilt-aversion?. *Journal of Economic Psychology*, **30**(5), 814–827.

Harth, N. S. and T. Regner (2009). Guilty Into Anger? Social Interaction in a Repeated Trust Game. *mimeo*

Ketelaar, T., and Au, W. T. (2003). The effects of guilty feelings on the behavior of uncooperative individuals in repeated social bargaining games: An Affect-as-information interpretation of the role of emotion in social interaction. *Cognition and Emotion*, **17**, 429–453.

Kube, S., Marechal, M. A., and Puppe, C. (2006). Putting Reciprocity to Work. Positive versus Negative Reciprocity in the Field. *University of St. Gallen working paper*, **2006**, 27.

Kugler, K. and Jones, W. H. (1992). On Conceptualizing and Assessing Guilt. *Journal of Personality and Social Psychology*, **62**, 318–327.

Manski, C. F. (2004). Measuring Expectations. *Econometrica*, **72**, 1329–1376.

McCabe, K.A., Rigdon, M.L., Smith, V.L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior and Organization*, **52**, 267–275.

Miettinen, T. and Suetens, S. (2008). Communication and Guilt in a Prisoner's Dilemma. *Journal of Conflict Resolution*, **52**, 945–960.

Moulton, R. W., Bernstein, E., Liberty, P. G. and Altucher, N. (1966). Patterning of parental affection and disciplinary dominance as a determinant of guilt and sex typing. *Journal of Personality and Social Psychology*, **4**, 356–363.

Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review*, **46**, 1423–1437.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, **83**(5), 1281–1302.

Reuben, E., Sapienza, P. and Zingales, L. (2009). Is mistrust self-fulfilling?. *Economics Letters*, **104** (2), 89–91.

Ross, L., D. Greene, and P. House (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, **13**, 279–301.

Ruffle, B.J. (1999). Gift giving with emotions. *Journal of Economic Behavior and Organization*, **39**, 399–420.

Segal, U., and Sobel, J. (2007). Tit for tat: Foundations for preferences for reciprocity in strategic settings. *Journal of Economic Theory*, **136**, 197–216.

Schotter, A. and Sopher, B. (2007). Advice and behavior in intergenerational ultimatum games: An experimental approach. *Games and Economic Behavior*, **58**, 365–93.

Tangney, J. (1995). Recent Advances in the Empirical Study of Shame and Guilt. *American Behavioral Science*, **38**, 1132–1145.

Trivers, R., (1971). The evolution of reciprocal altruism. *Quart. Rev. Biol.*, **46**, 35–57.

Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations. *Econometrica*, **76**(6), 1467–1480.

Zeelenberg, M., van Dijk, W. W., Manstead, A. S., and van der Pligt, J. (2000). On bad decisions and disconfirmed expectancies: The psychology of regret and disappointment. *Cognition & Emotion,*, **14** (4), 521–541.

# 6 Figures

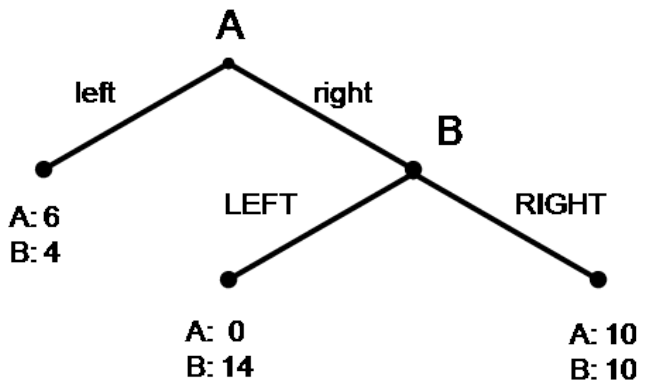



Figure 1: Structure of Game 1

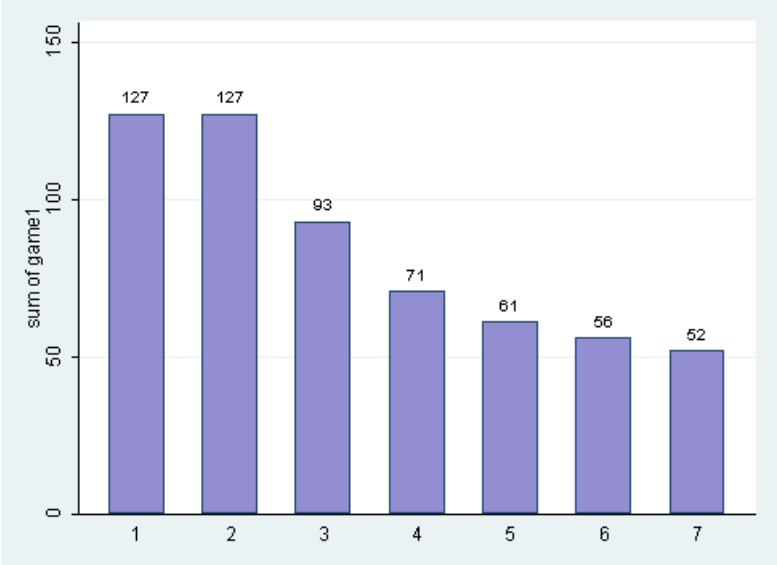
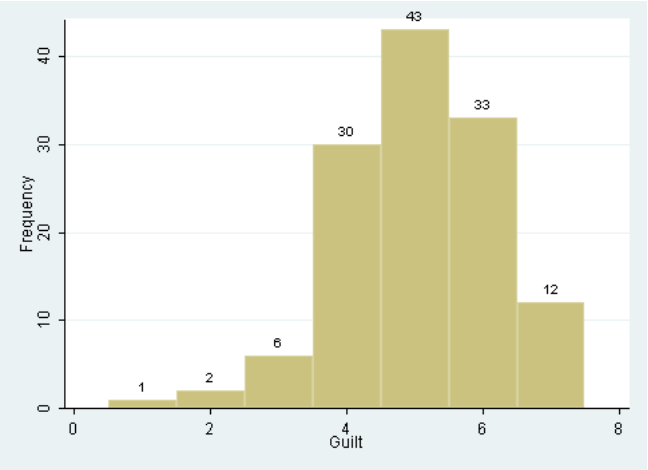
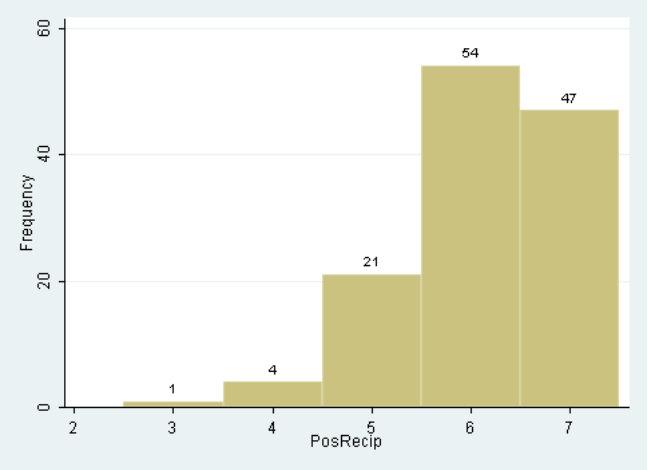


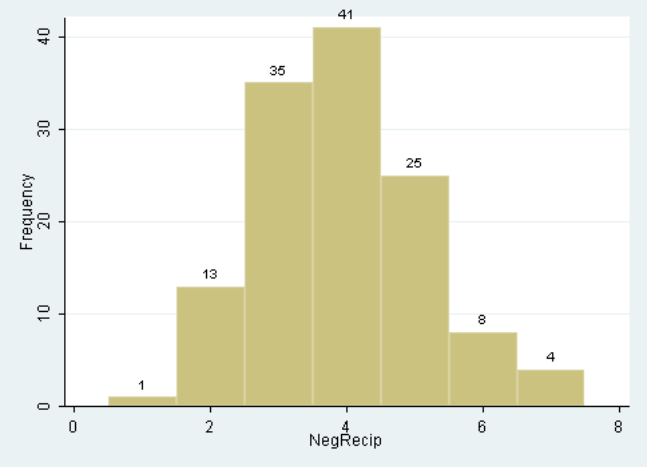

Figure 2: Number of participants playing Game 1 in each period. This number minus 127 equals the number of participants who played Game 2 in the respective period.




(a) Guilt




(b) Positive Reciprocity




(c) Negative Reciprocity

Figure 3: Self-assessed sensitivities to ...

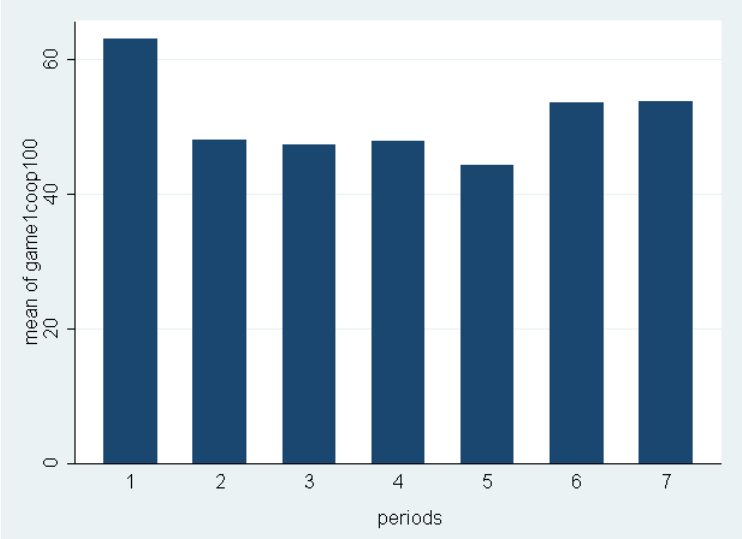



Figure 4: Cooperation rate in Game 1 over periods

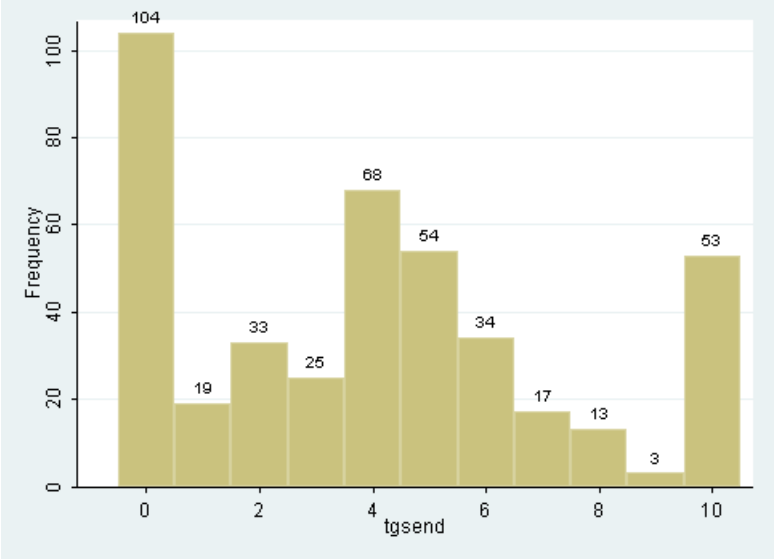



Figure 5: Histogram of amount sent in Game 2

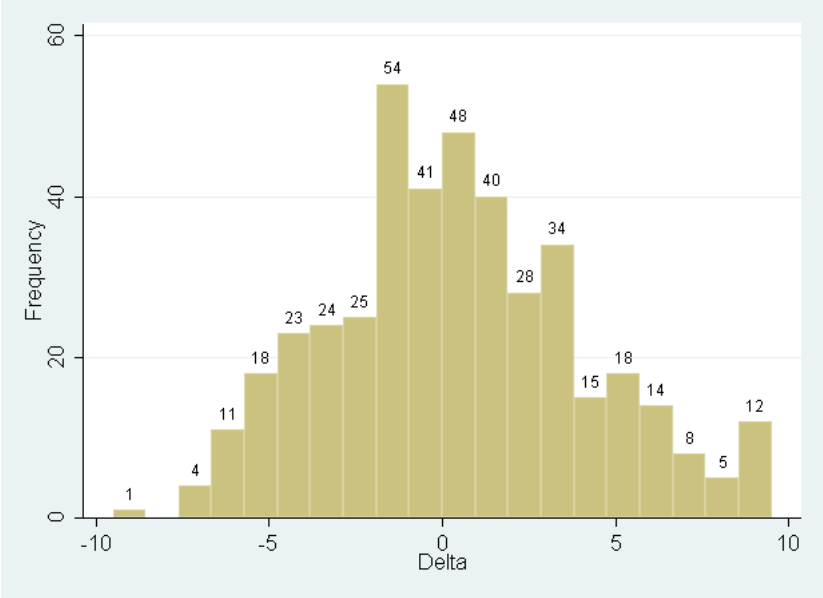



Figure 6: Histogram of Delta (the difference between the amount sent from $A$ to $B$ and the first-order belief of $B$) in Game 2

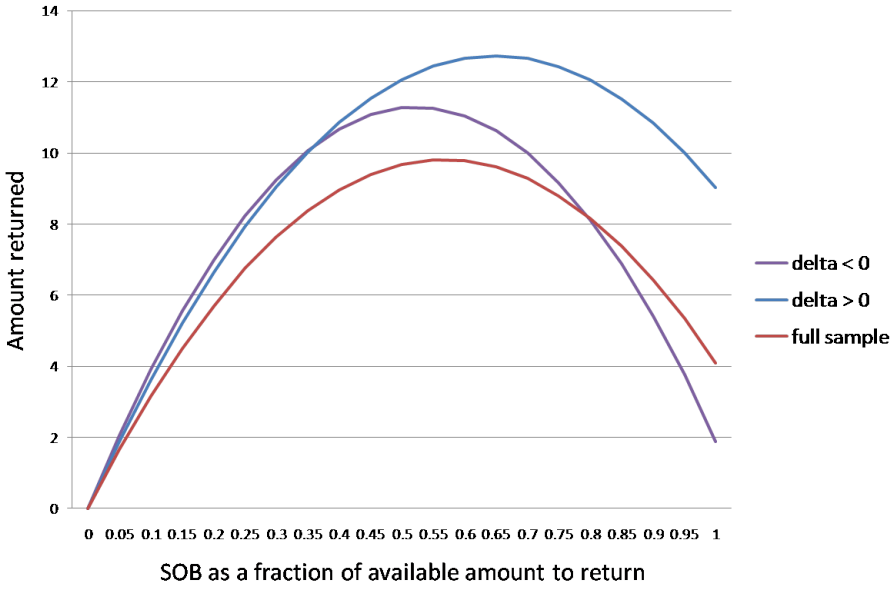



Figure 7: Estimates of the combined model in Game 2

# Appendix

## A. Questions used to assess general dispositions

**Ease with which guilt is experienced**

How easy is it for something to make you feel guilty?

**Positive reciprocity**

- If someone does me a favour, I am prepared to return it.
- I am ready to undergo personal costs to help somebody who helped me before.

**Negative reciprocity**

- If I suffer a serious wrong, I will take revenge as soon as possible, no matter what the cost.
- If somebody puts me in a difficult position, I will do the same to him/her.

## B. Experimental Instructions

Welcome and thanks for participating in this experiment.

In this experiment you can win a certain amount of money, which depends on your and the other participants' decisions in the experiment. **It is, therefore, important that you read the following instructions carefully.**

Please note that these instructions are only meant for you and that you are not allowed to exchange any information with the other participants. Similarly, during the entire experiment it is not allowed to talk to the other participants. If you have any questions or concerns, please raise your hand. We will answer your questions individually. Please do not ask your question(s) aloud. It is very important that you follow these rules, since otherwise we have to stop the experiment. Please also turn of your mobile phones now.

**General Procedure**

The experiment lasts about 60 minutes. Each decision will be explained again briefly on the monitor. While you make decisions, the other participants also make decisions which may influence your payoff.
During the experiment you can earn money. Your payoff will be calculated in ECU (Experimental Currency Units), 1 ECU = 0,75 EURO. At the end of the experiment your earnings will be converted into EURO and you are paid in cash. In addition you receive 2.50 Euro as a show-up fee.
Your payoff from the experiment depends on your decisions and the decisions of the other participants. But only one of the eight parts is chosen randomly and you are paid in cash according to the payoff from this part.
The exact procedure according to which your payoff is calculated is explained below. After you filled in a questionnaire the experiment ends and you receive your payoff.

Again the procedure as an overview:

1. Reading of the instructions, test questions (at the end of the instructions)
2. Decision situations
3. Questionnaire
4. Payoff and end of the experiment

**Details of procedure**

This experiment consists of eight parts in which two participants interact. They are called participant A and participant B.

Whether you are participant A or participant B will be determined randomly at the beginning of the experiment and you will stay in this role during the whole

experiment. Therefore it is very important that you **familiarise yourself with both roles**.

In each of these parts you are randomly and anonymously matched with another participants of the experiment.

The experiment consists of two parts in which you make decisions (Game 1 and Game 2). At the beginning of each part it will be determined, which of the two decision situations you will play. Therefore it is very important that you are **familiar with both decision situations**.

**Decision Situations**

**Game 1:**

In this game participant A will make a decision first. He/She can decide in favour of opportunity "left" or "right".

- The choice of "left" implicates a direct payoff, for example 6 ECU for participant A and 4 ECU for participant B.
- If participant A chooses opportunity "right", the payoffs of both participants will be determined by participant B.

B can choose between two options:

- A decision of "LEFT" means a payoff of 0 for participant A and a payoff of 14 ECU for participant B.
- A decision of "RIGHT" means a payoff of 10 ECU for participant A and a payoff of 10 ECU for participant B.

Participant B will always be asked for his/her decision, regardless if participant A has chosen "left" or "right".

It is not possible to observe the decisions of the other participant in game 1.

The following chart illustrates game 1 and the payoffs which arise:

See Figure 1

Please note that payoffs can change from round to round. **However, the structure of the game always remains constant.** If the payoffs change we will inform you individually.

**Game 2:**

In this game both participants have an endowment of S = 10 ECU.

First, participant A makes a decision. He/She can send an amount y (between 0 and the endowment S) to participant B. This amount will be doubled and placed at participant B's disposal. Hereupon participant B can decide how much he/she wants to send back to participant B. The amount which is sent back, z, can range from 0 to S + 2 * y. Both participants will be informed about y as well as z at the end of the round.

Consequently the following payoffs arise:

- Participant A: S - y + Z
- Participant B: S + 2 * y - z

[A chart illustrating game 2 and the payoffs which arise]

**Estimation**

Besides the choice of your actions you will be asked for

- your expectation concerning the other participant's action
- your estimation of the expectation of the other participant concerning your own action

You can earn money with these estimates. **The closer you are to the real amount the more you earn.**

**Expectations of participant B**

During the interaction as participant B you will be asked about your expectations concerning the decisions of participant A. Additionally you will be asked to give an estimation concerning the expectation of participant A concerning your own decision.

You are able to split your estimation in different intervals. Please indicate the estimated probability with a value between 0 and 100. Please consider that all probabilities must result in a sum of 100.

**Earnings with estimations**

Your earning from these estimations depends on how close your estimations are to the observed values in the experiment. The closer they are to the real value, the more you earn.
The maximal earning per estimation is 4 ECU. The real value is, as far as possible, defined by considering all participants. In either case it will be optimal for you, to indicate your real estimations. On request, you can see (after the experiment) how your earning from the estimations has been calculated in detail.

**Your payoff from the experiment**

Your payoff consists of several components. Your earning in any particular round is calculated as presented above (decision situation plus estimation). For the payoff only the earnings in of the eight rounds is relevant. This part is chosen randomly at the end of the experiment. The according earning will be paid in cash to you directly after the end of the experiment, that means after you completed the following questionnaire.