Christina Strassmair:

Can intentions spoil the kindness of a gift? - An experimental study

Munich Discussion Paper No. 2009-4

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Can intentions spoil the kindness of a gift? - An experimental study*

Christina Strassmair
University of Munich†

March 19, 2009

Abstract

Consider a situation where person A undertakes a costly action that benefits person B. This behavior seems altruistic. However, if A expects a reward in return from B, then A's action may be motivated by the expected rewards rather than by pure altruism. The question we address in this experimental study is how B reacts to the intentions of A. We vary the probability, with which the second mover in a trust game can reciprocate, and analyze effects on second mover behavior. Our results suggest that the perceived kindness and its rewards are not spoiled by expected rewards.

Keywords: social preferences, intentions, beliefs, psychological game theory, experiment

JEL classification: D02, C91, D64

*I thank Klaus Abbink, Georg Gebhardt, Martin Kocher, Sandra Ludwig, Theo Offerman, Klaus Schmidt, Joep Sonnemans, Matthias Sutter, Eva van den Broek, Frans van Winden, Peter Wakker, participants of the CREED seminar at the University of Amsterdam and the Theory Workshop at University of Munich for helpful comments on earlier versions of this paper. Financial support from ENABLE *Marie Curie Research Training Network*, funded under the 6^{th} Framework Program of the European Union, is gratefully acknowledged. I kindly thank the Center for Experimental Economics of the University of Innsbruck for providing laboratory resources and CREED of the University of Amsterdam for the generous hospitality.

†*Affiliation:* University of Munich, Seminar for Economic Theory, Ludwigstraße 28 (Rgb.), 80539 Munich, Germany, email: christina.strassmair[at]lrz.uni-muenchen.de.

1 Introduction

Consider a situation where person A undertakes a costly action that benefits person B. This behavior seems altruistic. However, if person A expects a reward in return, e.g. from person B, then person A's action may be motivated by the expected rewards rather than by pure altruism. If the expected rewards are sufficiently high, even selfish individuals have an incentive to behave in this way. The question we address in this paper is how person B reacts to the intentions of person A. Does person B perceive person A's action as less kind if he expects person A to expect rewards, and - if person B can reciprocate - does he return less?

There are many situations where behavior seems altruistic but is obviously strategic. Companies, for example, give Christmas gifts to their business partners in order to improve the business relationship, hoping that this pays off in future transactions. Their business partners may well understand that the given Christmas gifts are part of the company's profit maximizing investment strategy. The question, however, is whether this knowledge spoils the kindness of the gifts and makes them less effective.

We address this question experimentally in a series of modified trust games. In these games we vary the probability, with which the second mover can reciprocate, and investigate effects on second mover behavior. Our results suggest that neither the perceived kindness of the first mover's action nor the second mover's rewards are spoiled by expected future rewards.

In our modified trust game agent A, the first mover, decides how much of his initial endowment he transfers to agent B, the second mover. Then, a lottery determines whether agent B can decide on his return transfer to agent A or not. If not, nothing is returned to agent A. We conduct two treatments of this modified trust game that differ in the probability, with which agent B can decide on his return transfer: In treatment T-HIGH this probability is 80 % and in treatment T-LOW it is 50 %.[1] Agent A behaves in a way that seems altruistic when he transfers a strictly positive amount to agent B. This is true in both treatments. Our treatment variation, however, changes the possibility for agent B to make a return transfer to agent A and thereby, reduces the chance for agent A to receive a return. Hence, agent A's expected returns for his transfer are smaller in T-LOW when agent A has the same belief about agent B's reaction in both treatments. Consequently, agent B may perceive agent A as more kind in T-LOW than in T-HIGH and therefore,

[1] We do not implement probabilities close or equal to 1 and close or equal to 0 since we would like to avoid the effect that a certain event is going to happen almost for sure.

may return more in T-LOW than in T-HIGH when he is asked to decide. Agent A's beliefs about agent B's reaction, however, may differ in both treatments. We show that models of intention-based reciprocity predict that agent B returns (weakly) more in T-LOW than in T-HIGH. Nevertheless, agent A expects (weakly) smaller future rewards for a given transfer in T-LOW than in T-HIGH. This is because the difference in the probability, with which agent B can decide on his return transfer, dominates the difference in agent B's equilibrium return transfer.

Our results suggest that expected future returns do not affect the perceived kindness of an action and its rewards. Agent B's return transfer does not differ across treatments for a given transfer.[2] This is not because agent B does not care about agent A's action at all. Actually, we observe a lot of agents B that return strictly positive amounts, and in addition, agent B's average return transfer is increasing in agent A's transfer. This suggests that individuals reward actions that seem altruistic, independent of the actor's expectation of future rewards or of the actor's specific kind of intention. Consequently, we conclude that individuals condition their behavior on outcomes rather than on intentions or higher order beliefs.

We try to explain our findings by analyzing data from our questionnaire that each participant filled out after all decisions in a session had been made. First, our regressions of agent B's return transfer on agent B's (possibly incorrect) second order belief give no indication that expected future returns spoil the kindness of an action and its rewards. Second, we analyze whether agent B's perception of agent A's action is affected by the treatment. We do not find a significant effect. Third, we test treatment differences in agent B's stated emotions. For some interior values of agent A's transfer negative emotions like anger and contempt are experienced significantly more intense in T-HIGH than in T-LOW, while appreciation is significantly more pronounced in T-LOW than in T-HIGH. Even though intentions may affect individuals' emotions, these effects do not seem to carry over to the perception of an action and the reaction to it.

Intentions have been modelled in a number of theoretical papers and have been experimentally examined. Rabin (1993), Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006) are examples of models of intention-based reciprocity

[2]Furthermore, agent B's average return transfer in our treatments is not higher than the one in the trust game by Berg et al. (1995), in which the probability, with which that the second mover can decide on his return transfer, is 100 %. Note, however, that in the trust game by Berg et al. (1995) the second mover has the same intial endowment as the first mover. This fact may change the second mover's decision problem.

that take into account that intentions and higher order beliefs affect the perception of others' actions and, thereby, behavior. Blount (1995), Offerman (2002), Falk et al. (2003), McCabe et al. (2003), Charness (2004), Cox (2004) and Falk et al. (2008) experimentally test whether the second mover's reaction to the first mover's decision systematically differ when the first mover's decision is intentional rather than non-intentional, i.e. when the first mover has at least two choices available at his decision node rather than only one. The results of these studies on intentionality are mixed and depend on the experimental game that is implemented. In our study, in contrast, agent A's decision is intentional in all treatments[3] but the specific kind of intentions differs across treatments.

Stanca et al. (2009) analyze in their experimental study whether the second mover's reaction differs when the first mover's action is extrinsically motivated rather than intrinsically.[4] They hypothesize, and also find, that the slope of the second mover's reaction function is larger when the first mover is intrinsically motivated. In our experimental study, in contrast, we do not distinguish between extrinsic and intrinsic motivation since the first mover may expect a strictly positive return in both treatments and, therefore, may be extrinsically motivated in both treatments. We present models of intention-based reciprocity that predict the second mover to return more *for a given transfer* in T-LOW than in T-HIGH.[5]

The paper proceeds as follows. Section 2 presents the experimental design and procedure, Section 3 the behavioral predictions and hypotheses. Our results are summarized and discussed in Section 4. Section 5 concludes.

2 Experimental design and procedure

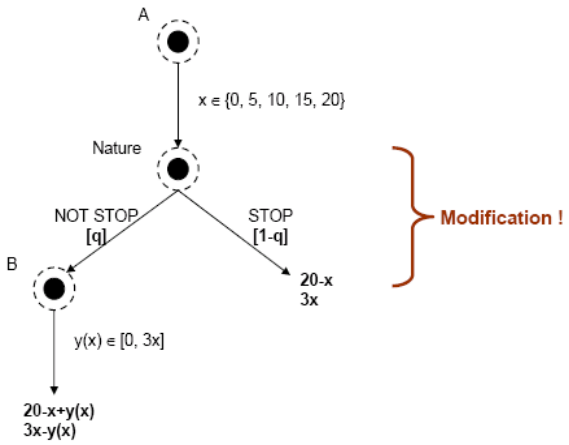
We consider a modified trust game with two agents, A and B. Agent A, the trustor, is initially endowed with $w_A = 20$ and can transfer an amount $x \in \{0, 5, 10, 15, 20\}$ to agent B, the trustee, who is initially endowed with $w_B = 0$. Agent B receives the tripled amount of agent A's transfer, $3 * x$. After agent A's decision on x a

[3] Moreover, agent A's set of actions does not vary across treatments.

[4] They compare the behavior of second movers in a standard trust game with the behavior of second movers in a trust game in which first movers are not informed that second movers can react on their transfer until they have made their decision. Hence, they implement an asymmetry of information conditions, which is not present in our experiment. In our experiment all participants (in all treatments) receive all relevant information at the beginning of the experiment.

[5] This prediction does not necessarily imply that the slope of the second mover's reaction function is larger in T-LOW than in T-HIGH.

Figure 1: Structure of the game



lottery determines whether the game stops at this point in time or continues. With probability $1 - q$ the game stops and agent A earns his initial endowment minus his transfer, $20 - x$, while agent B earns agent A's tripled transfer, $3 * x$. With probability q, though, the game continues and agent B can transfer an amount $y(x) \in [0, 3 * x]$ to agent A. Then, agent A earns his initial endowment minus his transfer plus agent B's return transfer, $20 - x + y(x)$, and agent B earns agent A's tripled transfer minus his return transfer, $3 * x - y(x)$.[6] The structure of this game is summarized in Figure 1.

The modification of the trust game consists in the random move of nature after agent A's decision on x. If $q = 1$ the game resembles the standard trust game introduced by Berg et al. (1995)[7]. In contrast, if $q = 0$ the game boils down to a dictator game[8] in which agent B can never return anything to agent A. The higher $q \in (0, 1)$, the higher the chance that agent B can make a return transfer (given $x > 0$) and the more similar the game is to the standard trust game. The smaller $q \in (0, 1)$, the smaller the chance that agent B can make a return transfer (given $x > 0$) and the more similar the game is to a dictator game.

As we address the question whether the perceived kindness of an action that

[6] Note that agent A does not receive the tripled amount of agent B's return transfer.

[7] One major difference to the game introduced by Berg et al. (1995) is that in their version of the trust game agent B has also a strictly positive initial endowment (which equals w_A).

[8] One major difference to standard dictator games is that in the typical versions the dictator's transfer is not tripled.

seems altruistic and its rewards are reduced by the actor's expectation of future rewards, we vary q, the probability with which agent B can return a positive amount (given $x > 0$), across treatments and keep everything else constant. Table 1 presents our treatments.

Table 1: Treatments

Treatment	q	Number of participants
T-HIGH	0.8	40
T-LOW	0.5	60

In treatment T-HIGH q is higher than in treatment T-LOW. We do not implement probabilities close or equal to 1 and close or equal to 0 since we would like to avoid the effect that a certain event is going to happen almost for sure. Furthermore, we are restrained to take higher values of q since agents B are only asked to decide on $y(x)$ when the game, indeed, continues[9]. Hence, if q was small, we expected very few observations of $y(x)$ for a given number of participants.[10]

In each each experimental session one treatment of the modified trust game is conducted. The implemented treatment of a session is played once. At the beginning of each session the roles of the game are assigned randomly. Participants are informed about their assigned roles after they have correctly answered a set of control questions. Agents A are always asked to decide on x, while agents B are only asked to decide on $y(x)$ when the game continues after agent A's decision on x. Given agents B are asked to decide we elicit $y(x)$ by the strategy method, i.e. agents B are informed that the game continues but are not informed about x and decide on their return transfer for each possible x.[11] After all participants have made their decisions they fill out a questionnaire concerning their emotions, beliefs, perception of the other player's action and individual data such as gender, age and subject of

[9] We could have asked all agents B to decide on $y(x)$ *given the game continues*. Then, however, treatment effects may have also been caused by social preferences based on expected outcomes and it would have been difficult to disentangle the source of observed treatment effects.

[10] For instance, if $q = 0.2$ and 100 individuals participated in this treatment, 50 individuals were agents B out of which we expected 10 to be asked to make a decision on $y(x)$.

[11] We apply the strategy method here in order to get agent B's reaction function. We are aware that this elicitation method may affect $y(x)$. However, we expect this effect to be orthogonal to our treatment variation. Furthermore, Stanca et al. (2009) argue that the strategy method applied in their trust games does not significantly affect decisions.

studies.

Our experimental sessions were run in April 2008 at the Center for Experimental Economics of the University of Innsbruck, Austria. 100 individuals participated in the experiment which was conducted with the software z-Tree by Fischbacher (2007). Individuals were randomly assigned to sessions and could take part only once. The sessions were framed neutrally[12] and lasted about an hour. Subjects earned on average 10.34 €[13] including a show-up fee of 5 €.

3 Behavioral predictions and hypotheses

We address the question whether the perceived kindness of an action that seems altruistic, i.e. a costly action that benefits others, and its rewards are reduced by the actor's expectation to receive future rewards. This may be the case since on the one hand, future rewards can partially cover the actor's initial costs and on the other hand, they reduce the others' net benefit. In the presented modified trust game, for instance, agent A behaves in a way that seems altruistic when he transfers a strictly positive amount to agent B. First, agent A's strictly positive amount is costly because this amount is deducted from his initial endowment. Second, agent A's strictly positive amount benefits agent B since the tripled transfer is assigned to agent B. This is true for both treatments. Agent A's expectation to receive a transfer from agent B in return to his transfer may reduce the perceived kindness of agent A's action, namely to transfer a strictly positive amount. In particular, the more agent A expects in return for his (given) transfer, the more of agent A's initial costs are covered in expectation, the less expected payoff is assigned to agent B, and therefore, the less kind agent B may perceive agent A and the less agent B may, in fact, return when he is asked to decide. Our treatment variation changes the possibility for agent B to make a return transfer to agent A and reduces the chance for agent A to receive a return. Hence, agent A's expected returns for a given transfer are smaller in T-LOW when agent A has the same belief about agent B's reaction in both treatments. Consequently, agent B may perceive agent A as more kind in T-LOW and therefore, may return more in T-LOW when he is asked to decide. If this, indeed, is the case and agent A's belief about agent B's reaction is correct and agent B's belief about agent A's belief is correct, then agent A expects agent B to transfer more in T-LOW when agent B is asked to decide. Nevertheless, agent A,

[12] Translated instructions can be found in the appendix.

[13] The maximum payoff paid out was 23 € and the minimum payoff 5 €.

in this case, faces less expected future returns in T-LOW since the difference in q compensates the difference in agent B's reaction. If it did not and agent A expected higher future returns in T-LOW agent B perceived agent A as less kind in T-LOW and therefore, he transferred less in T-LOW. Consequently, agent A's expectation were incorrect.

In the following we present standard models of social preferences that differ in their assumptions on individuals' utility function and, consequently, in their behavioral predictions. Some of them explicitly model how the perceived kindness is reduced by the actor's expectation to receive future rewards and predict that agent B returns more in T-LOW for a given transfer.

3.1 Behavioral predictions

3.1.1 Model 1: The self-interest model

The standard neoclassical model assumes that all individuals are selfish, i.e. their utility function U depends on their own material payoff m only and is increasing in m.

Given these assumptions agent B's decision does not vary in $q \in (0, 1)$.

As agent B maximizes his own material payoff only, he transfers $y(x) = 0 \; \forall \; x$ in the unique subgame perfect Nash equilibrium. This is true for all $q \in (0, 1)$.

3.1.2 Model 2: A model of social preferences based on final outcomes

Models of social preferences based on final outcomes such as e.g. those by Andreoni and Miller (2002), Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) assume that an individual's utility function $\widetilde{U}$ does not only depend on m but also on another individual's material payoff r. This does not necessarily imply that an individual is altruistic. Individuals with $\widetilde{U}$ may also be spiteful, envious, inequity averse or inequity loving.

Given these assumptions agent B's decision does not vary in $q \in (0, 1)$.

As agent B's decision is affected by final outcomes only[14] (and not e.g. by how these outcomes came about), agent B faces the same decision problem at his decision node independent of $q \in (0, 1)$. Hence, his optimal decision does not vary across treatments.

[14] Models of social preferences based on expected outcomes such as e.g. the one by Trautmann (2009) predict the same when agent B's decision is based on his expectations formed at the moment of his decision making.

3.1.3 Model 3: Models of intention-based reciprocity

Models of intention-based reciprocity such as e.g. those by Rabin (1993), Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006) assume that an individual's utility function V is not only dependent on outcomes but also on how these outcomes came about, e.g. whether the underlying decision problem was determined randomly or whether the underlying decision problem was intentionally brought about by another individual. A crucial role plays the perceived kindness of an individual's own action and the perceived kindness of other individuals' actions. Typically, the kinder an individual perceives the action of another individual, the kinder the individual treats this other individual. The perceived kindness of an action is shaped by the actor's intentions. How kindness is defined exactly and how intentions concretely enter the utility varies across models. In the following we present the predictions of a modified version of the model by Dufwenberg and Kirchsteiger (2004) and a similar model that implements central elements from the model by Falk and Fischbacher (2006).[15]

A modified version of the model by Dufwenberg and Kirchsteiger (2004)

As the model of Dufwenberg and Kirchsteiger (2004) is intended for finite multistage games without nature, we take their model and modify it in a simple and straight-forward way that accounts for random moves of nature. Our modification consists in the way how agent B perceives the kindness of agent A's strategy in the course of the game, in particular after the lottery determined that the game continues.[16]

Given the assumptions of this model $y(x)$ is (weakly)[17] higher in T-LOW than in T-HIGH for $x = 20$ in any sequential reciprocity equilibrium (SRE), in which agent B chooses a pure strategy.

A modified version of the model by Dufwenberg and Kirchsteiger (2004) with central elements of the model by Falk and Fischbacher (2006)

We take our modified version of the model by Dufwenberg and Kirchsteiger (2004) and implement central elements of the model by Falk and Fischbacher (2006)

[15] In the appendix we present these models and derive their predictions.

[16] We discuss the details of this modification in the appendix.

[17] No treatment differences are predicted if either agent B is hardly sensitive to reciprocity concerns such that he chooses $y(x) = 0$ in both treatments, or agent B is extremely sensitive to reciprocity concerns such that he chooses $y(x) = 3 * x$ in both treatments.

that concern how kindness is defined.[18]

Given the assumptions of this model $y(x)$ is (weakly)[19] higher in T-LOW than in T-HIGH $\forall\ x > 0$ in any sequential reciprocity equilibrium (SRE), in which agent B chooses a pure strategy.

3.2 Hypotheses

The various theoretical models predict different behavioral patterns of agent B. We focus on the predicted equilibria in which agent B chooses a pure strategy and summarize these predictions in the following three hypotheses.

Hypothesis 1: No returns in all treatments
Agent B returns nothing to agent A in T-HIGH and in T-LOW.

This hypothesis is supported by the self-interest model and implies that $y(x) = 0$ for all x and all treatments. Actions that seem altruistic are never rewarded.

Hypothesis 2: The same returns in all treatments
Agent B returns a weakly positive amount to agent A. Agent B's return transfer for a given x is the same in all treatments.

Models of social preferences support this hypothesis. Similar to the self-interest model there are no treatment effects. In contrast to the self-interest model, agent B returns a weakly positive amount to agent A. Actions that seem altruistic are rewarded, independent of the actor's intentions.

Hypothesis 3: Higher returns in T-LOW
Agent B returns a weakly positive amount to agent A. $y(x)$ is higher in T-LOW than in T-HIGH for $x > 0$.[20]

Models of intention-based reciprocity take into consideration how a decision problem came about and therefore, capture the effect of intentions. They predict that

[18] The details of this model are discussed in the appendix.

[19] No treatment differences are predicted if either agent B is hardly sensitive to reciprocity concerns such that he chooses $y(x) = 0$ in both treatments, or agent B is extremely sensitive to reciprocity concerns such that he chooses $y(x) = 3 * x$ in both treatments.

[20] This is supported by the modified version of the model by Dufwenberg and Kirchsteiger (2004) with central elements of the model by Falk and Fischbacher (2006). The modified version of the model by Dufwenberg and Kirchsteiger (2004) predicts $y(x)$ to be higher in T-LOW than in T-HIGH for $x = 20$, but not necessarily for all $x > 0$. The reason for the possibly different predictions is that the two models use different definitions of kindness.

the perceived kindness of an action that seems altruistic and its rewards are reduced by the expectation of future returns.

4 Results

We, first, summarize the descriptive results of our experiment and compare them with standard results from trust games and dictator games. In a next step, we test our hypotheses and analyze whether the perceived kindness of an action that seems altruistic and its rewards are reduced by the expectation of future rewards. Finally, we try to explain our findings with data from our questionnaire.

4.1 Summary statistics

4.1.1 Behavior of agent A

Table 2 presents the mean and the standard deviation of agent A's transfer in T-HIGH, T-LOW and both treatments together.

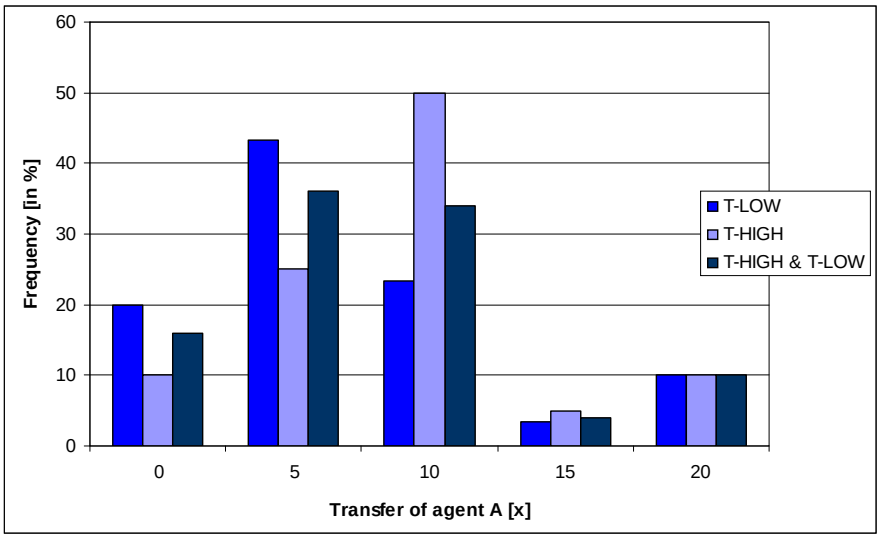
Table 2: Mean and standard deviation of x

Treatment	Mean	Standard deviation	Number of observations
T-HIGH	9.00	5.28	20
T-LOW	7.00	5.81	30
T-HIGH + T-LOW	7.80	5.64	50

On average, agent A transfers 7.8 points (out of 20 available points) to agent B. This is considerable larger than 0. Camerer (2003), however, reports that in standard trust games, in which $q = 1$ and agent B often has the same initial endowment as agent A, agent A on average transfers half of his initial endowment. This is relatively more than in our experiment, in particular in T-LOW, what suggests that agents A transfer less when the probability, with which agent B can reciprocate, is low. Figure 2 illustrates the aggregate distribution of x in T-HIGH, T-LOW and both treatments together.

In both treatments together 84 % of agents A transfer strictly positive amounts, more than 40 % half of their initial endowment or more and more than 10 % even

Figure 2: Distribution of x



more than 60 % of their initial endowment (some even their whole initial endowment). This is considerably different to results from standard dictator games, in which $q = 0$ and transfers are not tripled. For instance, in the benchmark treatment by Forsythe et al. (1994)[21] about 55 % of dictators transfer a strictly positive amount, less than 20 % half of their endowment or more and no dictator transfer more than 60 % of his endowment. This suggests that the distribution of x shifts towards higher values when $q > 0$ compared to $q = 0$.[22] When we consider the distributions of x in T-HIGH and in T-LOW, we observe that that the distribution is more to the right in T-HIGH than in T-LOW. Hence, it seems that agents A indeed react to differences in q. They tend to send more, the higher the probability, with which agent B can reciprocate.

4.1.2 Behavior of agent B

Table 3 presents the mean and the standard deviation of agent B's return transfer per x in T-HIGH, T-LOW and both treatments together.

In line with the results from the standard trust game by Berg et al. (1995), the more agent A transfers, the more agent B returns on average. The observed average return transfers, however, seem to be lower than the ones by Berg et al.

[21] Here, we refer to the paid dictator game conducted in April with a pie of 5 \$.

[22] This shift could also be caused by the fact that in standard dictator games agent A's transfer is not tripled. Cox (2004), though, observes that the distribution of transfers in a standard trust game is centred around higher values than the distribution of transfers in the corresponding trust game with $q = 0$.

Table 3: Mean and standard deviation of y(x)

Treatment	x	Mean	Standard deviation	y(x)/x	Number of observations
T-HIGH	5	01.75	02.59	0.35	16
	10	06.75	06.14	0.68	16
	15	11.06	08.83	0.74	16
	20	17.31	13.68	0.87	16
T-LOW	5	01.93	02.76	0.39	15
	10	04.93	04.35	0.49	15
	15	10.13	08.19	0.68	15
	20	16.47	12.00	0.82	15
T-HIGH + T-LOW	5	01.84	02.63	0.37	31
	10	05.87	05.34	0.59	31
	15	10.61	08.40	0.71	31
	20	16.90	12.69	0.85	31

(1995).[23] Table 3 also reports the rate of the average return transfer, i.e. the average return transfer divided by the transfer. Independent of $x > 0$ the rate of average return transfer is below 1. Hence, a strictly positive transfer does not pay on average fo agent A, even if agent A knew beforehand that the game is not stopped. Nevertheless, the rate of the average return transfer is increasing in x and peaks at a value more than 0.8 at $x = 20$.

If we separately examine agent B's return transfers in the two treatments, we observe that on average agent B returns more in T-HIGH than in T-LOW for all $x \in \{10, 15, 20\}$.

4.2 Analysis of hypotheses

4.2.1 Hypothesis 1: No returns in all treatments

Table 3 shows that agent B's average return transfers are considerably higher than 0 for $x > 0$. P-values of one sample median t-tests on $y(x) = 0$ per treatment and per x are reported in Table 4.

On the basis of these tests, we reject hypothesis 1 for all $x > 0$ and all treatments.

[23] One explanation for this difference could be that in the experiment by Berg et al. (1995) agents B have the same initial endowment as agents A.

Table 4: P-values of one sample median t-tests on hypothesis 1

Treatment	x	Percentage of agents B with y(x) = 0	Number of observations	p-value
T-HIGH	5	62.50	16	0.015
	10	31.25	16	0.001
	15	25.00	16	0.001
	20	25.00	16	0.001
T-LOW	5	60.00	15	0.015
	10	33.33	15	0.002
	15	20.00	15	0.001
	20	20.00	15	0.001

Nevertheless, Table 4 shows that there are some agents B that return nothing for some values of $x > 0$. The percentage of these observations decreases in x. Still, 25 % of agents B in T-HIGH and 20 % of agents B in T-LOW return nothing given $x = 20$.

4.2.2 Hypothesis 2: The same returns in all treatments

Table 3 indicates that agent B's average return transfer for a given $x > 0$ does not considerably vary across treatments. Table 5 reports per x the two-sided p-values of Mann-Whitney-U tests on whether $y(x)$ differs across treatments.

Table 5: Two-sided p-values of pairwise Mann-Whitney-U tests on hypothesis 2

x	Number of observations in T-HIGH	Number of observations in T-LOW	p-value
5	16	15	0.856
10	16	15	0.405
15	16	15	0.873
20	16	15	0.873

On the basis of these tests, we are far from rejecting hypothesis 2. Agent B's return transfer does not seem to differ across treatments.

4.2.3 Hypothesis 3: Higher returns in T-LOW

From the results presented in Table 5 we conclude that $y(x)$ is not significantly smaller in T-HIGH than in T-LOW neither for $x = 20$ nor for any other $x > 0$. If anything differed between T-HIGH and T-LOW regarding $y(x)$, then $y(x)$ was larger in T-HIGH than in T-LOW, at least for $x \in \{10, 15, 20\}$. Hence, our data seem to be inconsistent with hypothesis 3. One may, however, argue that the presented models of intention-based reciprocity predict no treatment difference either when agent B is hardly sensitive to reciprocity concerns such that he chooses $y(x) = 0$ in both treatments, or when agent B is extremely sensitive to reciprocity concerns such that he chooses $y(x) = 3 * x$ in both treatments. Table 4 reports that in both treatments a fraction of agents B return nothing even if $x = 20$. This suggests that a fraction of agents B are, indeed, hardly sensitive to reciprocity concerns. We, however, do not observe a single individual with $y(x) = 3 * x$ in any of our treatments. This rules out the possibility that a fraction of agents B are extremely sensitive to reciprocity concerns such that no treatment effects are predicted. Hence, either agents B are, in general, not sensitive to reciprocity concerns, or too few of agents B are sufficiently sensitive to reciprocity concerns.

We summarize our findings in the following two results:

Result 1: Rewards for actions that seem altruistic

As in previous studies on trust games (for an overview see Camerer, 2003) we observe that agent B returns significantly positive amounts. These amounts on average increase in agent A's transfer.

Result 2: No effect of the intention[24]

Agent B's return transfer (for a given $x > 0$) does not vary across treatments.

These results are consistent with the predictions by models of social preferences, but inconsistent with the predictions by the self-interest model. The presented models of intention-based reciprocity may predict no treatment differences, but only for individuals that are sufficiently insensitive to reciprocity concerns or for individuals that are extremely sensitive to reciprocity concerns. In our data, however, there is no evidence for individuals that are extremely sensitive to reciprocity concerns. There may be individuals that are sufficiently insensitive to reciprocity concerns and return nothing. On average, though, agents B return strictly positive amounts.

[24] Here, we do not question the effect of intentionality, though.

Consequently, the predictions of the presented models of intention-based reciprocity are inconsistent with our aggregate results. We conclude that the kindness of an action and its rewards are not spoiled by the actor's expectation to receive future rewards. On average, actions that seem altruistic are rewarded by others. The rewards vary in the action. The more altruistic they seem, the higher are the average rewards. The average rewards for a given action, though, are independent of the actor's expectation to receive future rewards.

4.3 Possible explanations for our findings

In this section we try to find explanations for our findings by analyzing data from our questionnaire.

4.3.1 Incorrect higher order beliefs of agent B

The perceived kindness of an action that seems altruistic can only be spoiled by the actor's expectation of future rewards, if individuals expect the actor to expect future rewards. From other experimental studies we know that individuals have difficulties to draw inferences from other individual's actions and correctly form beliefs.[25] In the following, we analyze whether the given elicited second order beliefs of agent B directly affect his behavior.[26] We regress agent B's return transfer for a given x on x and on the product of agent B's second order belief with q for a given x, i.e. agent B's expectation of agent A's expected future returns for a given x. First, we estimate OLS regressions with and without controls such as sex, age and subject of studies. Second, we run two two-stage least squares instrumental variable regressions, in which we instrument for the product of agent B's second order belief with q for a

[25] Prominent examples are experimental studies on information cascades, e.g. by Anderson and Holt (1997), Hung and Plott (2001), Kariv (2005), Nöth and Weber (2003) and Goeree et al. (2007).

[26] Agent B's second order belief was elicited in a non-incentivized way after agent B has made his decision. We are aware that these second order beliefs may be affected by agent B's own decision. Therefore, we checked whether agent B's elicited second order belief significantly differs from those elicited by agents B who have not decided upon $y(x)$ because the lottery stopped the game after agent A's decision. We run pairwise Mann-Whitney-U tests and do not find a significant difference. Hence, we assume that an agent's own action does not influence his second order beliefs to a large extent.

given x[27].[28] Table 6 presents the results of our four regressions for $x > 0$.[29]

Table 6: Regressions of the return transfer for x >0

Dependent variable: y(x)	OLS-c1	OLS-c2	2SLS-IV-c1	2SLS-IV-c2
Intercept	- 03.05***	- 01.44	- 01.67	+07.38
x	+00.79***	+00.71**	+00.33	- 00.70
Agent B's second order belief * q	+00.24	+00.33	+00.77	+01.96
Sex		- 03.08		- 06.62
(1 if male, 0 else)				
Age		+00.06		+00.00
Subject of studies		- 01.94		- 04.50
(1 if economist, 0 else)				
Number of observations	124	124	124	124
R-squared	0.3384	0.3697	0.2700	< 0

*, **, *** significant at 10, 5, 1 percent significance level
-c with individual clusters

In all of our regressions agent B's belief about agent A's expected return does not significantly affect agent B's return for a given x. In OLS-c1 and OLS-c2 the only significant regressor is agent A's transfer: The higher agent A's transfer, the more agent B returns. The estimated coefficients of our control variables are all insignificant.

Result 3: No effect of agent B's belief about agent A's expected returns
Agent B's elicited beliefs about agent A's expected returns do not affect agent B's returns.

Consequently, we conclude that incorrect higher order beliefs of agent B are not the explanation for why the kindness of an action that seems altruistic and its future rewards are not spoiled by the actor's expectation to receive future rewards.

[27] The instrument we use is q itself as it is exogenous and highly correlated with the instrumented variable.

[28] We run these two additional regressions since agent B's second order belief for x could be endogenous and, therefore, our estimated OLS coefficients could be biased and inconsistent.

[29] In all regressions we consider $x > 0$ since the restriction on $x = 20$ would considerably reduce our data set.

4.3.2 Effect only on the perception or on emotions

The perceived kindness of an action may be spoiled by the actor's expectation to receive future rewards without affecting the reactions to that action. In addition, the actor's expectation may affect the reactor's emotions in the sense that he experiences more negative emotions and less positive emotions in T-HIGH than in T-LOW. Table 7 reports one-sided p-values of Mann-Whitney-U tests on whether the perceived kindness of agent A's action[30] differs across treatments.

Table 7: One-sided p-values of Mann-Whitney-U tests on perceived kindness across treatments

x	Number of observations in T-HIGH	Number of max. int. in T-HIGH	Number of observations in T-LOW	Number of max. int. in T-LOW	p-value
0	16	0	15	0	0.1058
5	16	0	15	0	0.0964
10	16	0	15	1	0.0284
15	16	1	15	2	0.2567
20	16	10	15	12	0.2294

max. int. observations perceiving the other's kindness with maximal intensity
There are no observations that perceived the other's kindness with minimal intensity.

For any $x \in \{0, 15, 20\}$ we do not identify any significant differences in the perceived kindness across treatments.[31] $x = 5$ and $x = 10$ are perceived as less kind in T-HIGH at a significance level of 10 %. We take this as weak evidence that agent B's perception of agent A's action is affected by the treatment variation.

Result 4: Hardly no effect of the intention on agent B's perception of agent A's action

Agent B's perceived kindness of agent A's action does not significantly vary across treatments. This is true for $x = 20$ and for the most other values of x.

Table 8 reports one-sided p-values of Mann-Whitney-U tests on whether the intensity of hypothetically sensed emotions[32] is higher in one of the treatments.

[30] In our questionnaire agents B had to indicate on a scale ranging from 0 to 7 how kind they perceive a given transfer by agent A. 0 represented "very unkind", while 7 represented "very kind".

[31] For $x = 20$ this could be due to the fact that the majority of agents B choose the maximal intensity.

[32] In our questionnaire individuals had to indicate on a scale ranging from 0 to 7 with which

For $x = 20$, stated anger is sensed significantly more strongly in T-HIGH than in T-LOW at a significance level of 10 %. There is no significant difference in the sensed contempt, gladness and appreciation for $x = 20$.[33]

Table 8: One-sided p-values of Mann-Whitney-U tests on emotions across treatments

Emotion	x	Number of observations in T-HIGH	Number of max. int. in T-HIGH	Number of observations in T-LOW	Number of max. int. in T-LOW	p-value
Anger	0	16	3	15	1	0.1449
	5	16	0	15	0	0.0669
	10	16	0	15	0	0.0154
	15	16	0	15	0	0.0023
	20	16	0	15	0	0.0820
Contempt	0	16	4	15	0	0.0358
	5	16	0	15	0	0.0384
	10	16	0	15	0	0.0079
	15	16	0	15	0	0.0097
	20	16	1	15	0	0.4185
Gladness	0	16	0	15	0	0.2011
	5	16	1	15	0	0.1242
	10	16	1	15	0	0.0989
	15	16	1	15	3	0.3273
	20	16	12	15	13	0.1791
Appreciation	0	16	0	15	0	0.1439
	5	16	0	15	0	0.0054
	10	16	0	15	1	0.0060
	15	16	0	15	3	0.0579
	20	16	9	15	9	0.3954

max. int. observations sensing an emotion with maximal intensity
There are no observations that sensed an emotion with minimal intensity.

For interior values of x, anger is significantly more strongly pronounced in T-HIGH than in T-LOW. The same holds for contempt. Only for $x = 10$ gladness is significantly less pronounced in T-HIGH than in T-LOW at a significance level of

intensity they hypothetically sensed an emotion for each x. If they did not sense an emotion at all, they were asked to indicate 0 for this emotion and the given x.

[33] One may, however, argue that regarding the positive emotions a large number of observations indicated the maximal intensity and therefore, no treatment differences are identified.

10 %. Appreciation is significantly less pronounced in T-HIGH than in T-LOW for interior values of x.

For $x = 0$ we detect a significant treatment difference in contempt only.

Result 5: Effect of the intention on anger, contempt and appreciation for interior values of x

Negative emotions such as anger and contempt are significantly more strongly pronounced in T-HIGH than in T-LOW for interior values of x. Furthermore, appreciation is significantly less strongly pronounced in T-HIGH than in T-LOW for interior values of x. Gladness seems to be unaffected by the treatment variation for the most values of x.

Consequently, we conclude that agent B's emotions may be affected by agent A's intentions. This effect, however, does not seem to carry over to agent B's perception of agent A's action and to agent B's reaction.

4.3.3 Other explanations

There are other potential reasons for why the perceived kindness of an action that seems altruistic and its rewards are not spoiled by the actor's expectation of future rewards in our setting. One reason may be that agent B can voluntarily decide on his return transfer and is not forced to return a certain amount. Expecting a return that is voluntarily given may not spoil the kindness of an action. This may be different for expecting a return that is not voluntarily given. The presented models of intention-based reciprocity do not take this into account.

Another reason may be that kindness is not an absolute measure but a relative one that captures the ranking of actions for a *given* action set. $x = 20$, for instance, may be perceived as the kindest action of agent A and therefore, would be evaluated as equally kind in both treatments.

5 Conclusion

We have presented an experimental study on whether the perceived kindness of an action that seems altruistic, i.e. a costly action that benefits others, and its rewards are reduced by the actor's expectation to receive future rewards.

In our experimental study second movers in a modified trust game return significantly positive rewards to first movers. These rewards on average increase in

the first mover's transfer. They, however, do not significantly vary in the probability, with which the second mover can reciprocate. On average, the second mover's return transfer is even slightly higher when the probability that the second mover can reciprocate is 0.8 rather than 0.5 for some values of x. On the basis of data from our questionnaire we test whether this is due to incorrect higher order beliefs of second movers. Our results suggest that this, however, does not seem to be the case. Furthermore, we test whether the second mover's perception or emotions are influenced by the probability, with which the second mover can reciprocate. We find significant effects on some of the second mover's emotions, at least for some values of x.

Our results, therefore, suggest that behavior that seems altruistic is rewarded. The more altruistic it seems, the higher is the reward in return. The reward for a given action, however, does not vary in the actor's expectation to receive future rewards. This is consistent with the predictions of models of social preferences but inconsistent with the predictions of the self-interest model and the presented models of intention-based reciprocity. Hence, individuals in this setting seem to condition their behavior on outcomes rather than on intentions or higher order beliefs.

Our results seem to be relevant for different kinds of contexts. Political as well as commercial campaigns often try to gain the support of a large group of individuals by behaving in a way that seems altruistic, e.g. by distributing small gifts. Individuals may well anticipate that these gifts are intended to gain their support. In the light of our results, however, we would conclude that this does not diminish the effectiveness of the small gifts. Similarly, in some organizations workers are financially incentivized to help their colleagues.[34] Workers, therefore, may anticipate that the help of a colleague is motivated by receiving financial rewards. We would conclude that this does not diminish the perceived kindness of help and does not harm the willingness to reward this action.

This experimental study contributes to the discussion of higher order beliefs and of intentions. Our results suggest that higher order beliefs and specific kinds of intentions do not significantly influence behavior, at least the reaction to an action that seems altruistic. It may well be that higher order beliefs and intentions are crucial for other sorts of behavior, though, e.g. for the reaction to socially undesired behavior. Criminal law often conditions penalties on the criminal's intentions. Hence, the effect of intentions may depend on the specific context.

[34] A worker's wage may, for instance, be dependent on the performance of his colleagues.

6 Appendix

6.1 Experimental sessions and instructions

6.1.1 Experimental sessions

The order of events during each experimental session was the following: Individuals were welcomed and randomly assigned a cubicle in the laboratory where they took their decisions in complete anonymity from the other individuals. The random allocation to a cubicle also determined an individual's role. The instructions for the experiment, which each individual found in his cubicle, were read aloud. Then, individuals could go through the instructions on their own and ask questions. After all remaining questions were answered and no individual needed more time to go through the instructions, they had to answer a set of control questions concerning the procedure of the experiment. After each individual had answered all control questions correctly, they were informed about their role in the experiment and we proceeded to the decision stages. First, agents A decided upon x. Second, a computer programm determined randomly which games of a session were stopped. Each game in a session had the same probability that it is stopped, which corresponded to q of the implemented treatment of a session. Third, agents B were informed about whether game was stopped or not. In case the game was not stopped, agents B decided upon the return transfer for each x. In case the game was stopped, agents B were asked the question, what they would have transferred in return for each x if the game continued. During the course of the experiment individuals were asked questions whose answers were not related to any payments, e.g. agents A were asked after their decision on x how many points they believe agent B transfers in return for each x given the game is not stopped, and agents B were asked after their real or hypothetical decision on $y(x)$, respectively, which intensities of certain emotions they would experience for each x. After all participants answered the questions posed to them, all agents were informed about the outcome of the game, i.e. agent A's decision, nature's random move on whether the game stops right after agent A's decision, and - in case the game was not stopped - agent B's decision for the corresponding x. Finally, we elicited demographic variables such as gender, age and subject of studies. At the end of the session individuals were paid in cash according to their earned amount in the modified trust game plus a show-up fee of 5 Euro.

The instructions, the program, and the questionnaire were originally written in German. The translated instructions for T-HIGH can be found in the following.

The instructions for T-LOW are similar expect that the probability, with which the game is stopped right after agent A's decision, is $q = 0.5$.

6.1.2 Translated instructions of T-HIGH

Instructions for the experiment

Welcome to this experiment. You and the other participants are asked to make decisions. Your decisions as well as the decisions of the other participants determine the result of the experiment. At the end of the experiment you will be paid **in cash** according to the **actual** result of the experiment. So please read the instructions attentively and think about your decisions carefully. In addition, you receive – independent of the result of the experiment - a show up fee of 5 Euro.

During the whole experiment it is not allowed to talk with the other participants, to use mobile phones or to start other programs on the computer. The contempt of these rules immediately leads to the exclusion of the experiment and of all payments. If you have any questions, please raise your hand. An instructor of the experiment will then come to your seat in order to answer your questions.

During the experiment we talk about points rather than about Euros. Your whole income is initially calculated in points. At the end of the experiment your actual amount of total points is converted into Euros according to the following rate:

1 point = 30 Cents.

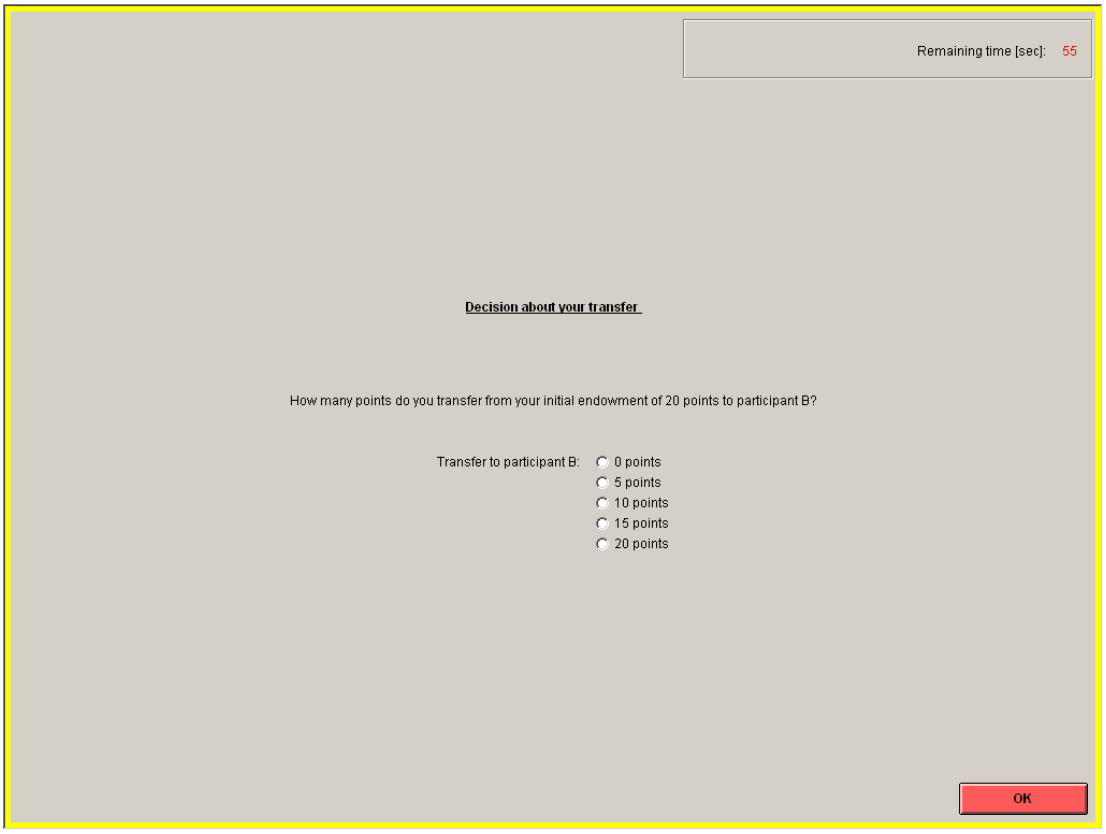
In this experiment, there are **participants A** and **participants B**. Before the experiment starts, you are informed whether you are a participant A or a participant B. While entering the room this was randomly determined. If you are participant A, you are randomly and anonymously matched to a participant B. If you are participant B, you are randomly and anonymously matched to a participant A. Neither during nor after the experiment, you receive any information about the identity of your matched participant. Likewise, your matched participant does not receive any information about your identity.

The procedure

Participant A has an initial endowment of 20 points. Participant B has an initial endowment of 0 points.

Participant A can decide how much of his initial endowment he transfers to participant B. **Participant A can either choose 0, 5, 10, 15 or 20 points**.

In order to make this decision, participant A selects one amount on the following computer screen and presses the OK-button.



Participant A's transfer is then **tripled** and sent to participant B.

After participant A chose his transfer and participant A's tripled transfer was sent to participant B, it is randomly determined, whether the experiment is stopped at this point in time.

- With the probability of 20% the experiment is stopped at this point in time. **In this case participant A receives his initial endowment minus his transfer, and participant B receives participant A's tripled transfer**.
- With the probability of 80% the experiment is not stopped at this point in time and participant B decides which integer between 0 and participant A's tripled transfer (including 0 and participant A's tripled transfer) he transfers back to participant A. **In this case participant A receives his initial endowment minus his transfer plus participant B's back transfer, and participant B receives participant A's tripled transfer minus his back transfer**.

In case the experiment is not stopped right after participant A's decision, participant B makes the decision about the back transfer. In order to do that participant

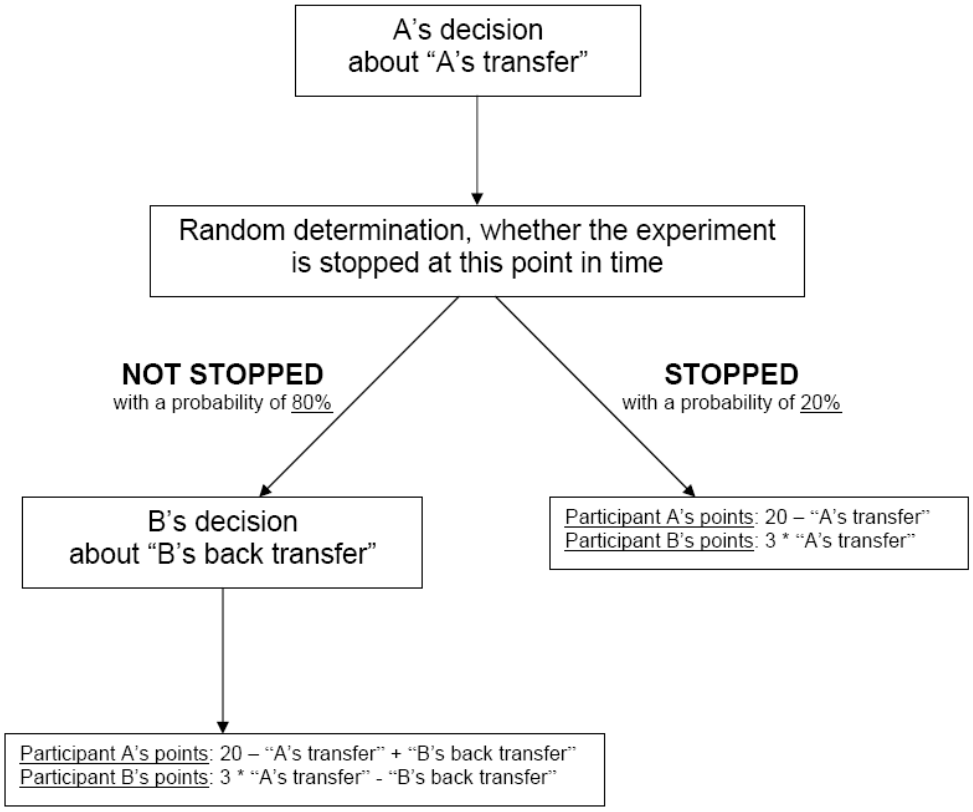
B indicates for each possible transfer of participant A his selected amount on the following computer screen and presses the OK-button. Depending on what participant A transferred, participant B's corresponding entry is transferred back to participant A.

Participant B makes this decision only if the experiment was not stopped right after participant A's decision.

Example 1: Participant A chooses a transfer of 15 points. Then, it is randomly determined that the experiment is stopped right after participant A's decision. Participant A receives 20 – 15 points = 5 points. Participant B receives 3 * 15 points = 45 points.

Example 2: Participant A chooses a transfer of 15 points. Then, it is randomly determined that the experiment is not stopped right after participant A's decision. Participant B chooses a back transfer of 39 points if participant A transferred 15 points. Participant A receives 20 – 15 + 39 points = 44 points. Participant B receives 3 * 15 – 39 points = 6 points.

The procedure is illustrated by the following graph:



After this procedure participant A and participant B are both informed about participant A's transfer, about whether the experiment was stopped right after participant A's decision, and - in case the experiment was not stopped right after participant A's decision - about participant B's back transfer. Then, the experiment ends. The procedure is not repeated.

During the course of the experiment you might be asked to answer questions. The answers to these questions do not affect the payments and the procedure of the experiment. They are treated anonymously and are not sent to your matched participant or any other participant.

Before you are informed whether you are participant A or participant B and the experiment starts, you are asked to answer several control questions concerning the procedure of the experiment.

If you have any questions, please raise your hand. An instructor of the experiment will then come to your seat in order to answer your questions.

6.2 Behavioral predictions of the modified version of the model by Dufwenberg and Kirchsteiger (2004)

6.2.1 The basic model by Dufwenberg and Kirchsteiger (2004)

In Dufwenberg and Kirchsteiger (2004) individual i's utility function in a 2-player game with individual j is defined in the following way:

$$U_i = \pi_i + Y_i * \kappa_i * \lambda_i,$$

where π_i represents individual i's expected own material payoff, $Y_i \geqslant 0$ individual i's parameter of sensitivity to reciprocity concerns, κ_i individual i's perception of the kindness of his own strategy and λ_i individual i's perception of the kindness of individual j's strategy. Y_i is a parameter that is exogenously given. π_i, κ_i and λ_i are dependent on individual i's strategy and on individual i's belief about individual j's strategy. Furthermore, λ_i also depends on individual i's belief about individual j's belief about individual i's strategy.

Dufwenberg and Kirchsteiger (2004) define κ_i as individual i's expectation of individual j's expected material payoff[35] minus a reference payoff, which is the mean of the maximum and the minimum expected material payoff individual i beliefs he could assign to individual j[36]. λ_i is defined as individual i's expectation of individual j's expectation of individual i's expected material payoff[37] minus a reference payoff, which is the mean of the maximum and the minimum expected material payoff individual i beliefs that individual j beliefs he could assign to individual i. In other words, λ_i is the kindness of individual j's strategy from the perspective of individual i.

[35] Individual i's expectation of individual j's expected material payoff may not be equal to π_j since individual i's belief about individual j's strategy may not be equal to individual j's strategy and individual i's belief about individual j's belief about individual i's strategy may not be equal to individual j's belief about individual i's strategy.

[36] Dufwenberg and Kirchsteiger (2004) define the reference payoff as the mean of the maximum and the minimum expected material payoff individual i beliefs he could assign to individual j given *efficient* strategies. They define an efficient strategy as a strategy for which there does not "exist another strategy which conditional on any history of play and subsequent choices by the others provides no lower material payoff for any player, and a higher material payoff for some player for some history of play and subsequent choices by others" (p.276). In our modified trust game all strategies are efficient and, therefore, we abstract from this more general definition.

[37] This may not be equal to individual i's expected payoff π_i, since individual i's belief about individual j's belief about individual i's strategy may not be equal to individual i's strategy.

Note that an individual's beliefs are updated in the course of the game and therefore, may differ after different histories of play.[38] Hence, an individual's perception of the kindness of his own strategy and of the other individual's strategy are updated in the course of the game and may differ after different histories of play. Dufwenberg and Kirchsteiger (2004) introduce the sequential reciprocity equilibrium (SRE), in which each player in each of his decision nodes makes choices that maximize his utility for the given history, given his updated first and second order beliefs and given that he follows his equilibrium strategy at other histories. Furthermore, all players' initial first and second order beliefs are correct and following each history are updated.

6.2.2 Our modification

Dufwenberg and Kirchsteiger (2004) restrict attention to finite multi-stage games without nature. However, we could simply consider nature as a third player who always chooses to stop the game with probability $1 - q$ and to continue the game with probability q, and to whom agent A and agent B are insensitive to reciprocity concerns. But this would lead to an unintuitive way of evaluating agent A's kindness in the course of the game: At the beginning of the game agent B has some belief about agent A's strategy and nature's strategy. After agent A's chosen amount is transferred and the lottery has chosen to continue the game, agent B's updated beliefs are that agent A chooses the given transfer (with probability 1), that nature chooses to continue the game (with probability 1) and that agent A beliefs that nature chooses to continue the game (with probability 1). If agent B evaluates the kindness of agent A's strategy given his updated beliefs, he takes into consideration that agent A beliefs that nature chooses to continue the game with probability 1. However, agent A's belief about nature's strategy was different at agent A's decision node and therefore, agent A's intentions were different.

In order that this is not the case, we undertake a small and natural modification of the basic model by Dufwenberg and Kirchsteiger (2004). Our modification consists in the way how agent B perceives the kindness of agent A's strategy in the course of the game. We let agent B evaluate the kindness of agent A's strategy given the belief that agent A beliefs that nature chooses to continue the game with probability

[38] For example, individual i may expect individual j to play action a with probability p and action b with probability $1 - p$ at the beginning of the game (which may, indeed, be correct). After individual j's action a has realized, individual i beliefs that individual j has chosen a with probability 1, and not p.

q rather than with probability 1.

6.2.3 Agent B's utility function when he is asked to decide

Consider agent A has chosen on x and the lottery has determined to continue the game. Agent B, then, decides on $y(x) \in [0, 3 * x]$ and beliefs that agent A has chosen x (with probability 1), that nature has chosen to continue the game (with probability 1) and that agent A beliefs that agent B returns $\widetilde{y}(0) = 0$, $\widetilde{y}(5) \in [0, 15]$, $\widetilde{y}(10) \in [0, 30]$, $\widetilde{y}(15) \in [0, 45]$, $\widetilde{y}(20) \in [0, 60]$, where $\widetilde{y}(x)$ represents agent B's second order belief for x.

Then, π_B, agent B's expected own material payoff, is equal to $3 * x - y(x)$ and κ_B, agent B's perception of the kindness of his own strategy $y(x)$, is equal to $(20 - x + y(x)) - \left(\frac{(20-x+0)+(20-x+3*x)}{2}\right)$ since the minimum he can assign to agent A is $20 - x + 0$ and the maximum $20 - x + 3 * x$. λ_B, agent B's perception of the kindness of agent A's strategy, is equal to $(3 * x - q * \widetilde{y}(x)) - ref$, where ref is the corresponding reference payoff that depends on agent B's second order beliefs[39].

6.2.4 Equilibrium predictions

In this subsection we derive some statements that hold in any SRE, in which agent B chooses a pure strategy.

Lemma 1a: $y(x)$ is (weakly) increasing in x in any SRE, in which agent B chooses a pure strategy, $\forall\ q \in (0, 1)$.

Suppose not. Then, there exists an $x' > x$ such that $y(x') < y(x)$ in SRE for some $q \in (0, 1)$. As in any SRE actions and beliefs about these actions are the same, agent B perceives agent A as more kind, when he receives $3 * x'$ than when he receives $3 * x$. This is true since agent A assigns more expected payoff to agent B, when he chooses x' instead of x:

$$3 * x' - y(x') * q > 3 * x - y(x) * q,$$

and faces the same reference payoff for x' and x. Hence, $\lambda_B(x') > \lambda_B(x)$. Nevertheless, agent B returns less when he receives $3 * x'$ than when he receives $3 * x$. In order that this is, indeed, optimal for agent B the following two (weak) inequalities are necessary:

[39] The maximum agent B beliefs agent A beliefs he can assign to agent B is not necessarily equal to $3 * 20 - q * \widetilde{y}(20)$.

$3*x'-y(x')+Y_B*\lambda_B(x')*\big(y(x')-\frac{3}{2}*x'\big) \geq 3*x'-y(x)+Y_B*\lambda_B(x')*\big(y(x)-\frac{3}{2}*x'\big)$,

because agent B could also send back $y(x) < 3*x'$ given x' but (weakly) prefers to send back $y(x')$ in this case, and

$3*x-y(x)+Y_B*\lambda_B(x)*\big(y(x)-\frac{3}{2}*x\big) \geq 3*x-y(x')+Y_B*\lambda_B(x)*\big(y(x')-\frac{3}{2}*x\big)$,

because agent B could also send back $y(x') < y(x)$ given x but (weakly) prefers to send back $y(x)$ in this case. The first (weak) inequality can be rewritten as

$$\tfrac{1}{Y_B} \geq \lambda_B(x'),$$

and the second as

$$\tfrac{1}{Y_B} \leq \lambda_B(x).$$

As $\lambda_B(x') > \lambda_B(x)$ this is a contradiction.

Lemma 1b: $\widetilde{y}(x)$ is (weakly) increasing in x in any SRE, in which agent B chooses a pure strategy, $\forall\ q \in (0,1)$.

As in any SRE, in which agent B chooses a pure strategy, actions and beliefs about these actions are the same, $y(x) = \widetilde{y}(x)\ \forall\ x$. Due to Lemma 1a $y(x)$ is increasing in x in any SRE and therefore, also $\widetilde{y}(x)$.

Lemma 2: $(3*x-y(x)*q)$ is (weakly) increasing in x in any SRE, in which agent B chooses a pure strategy, $\forall\ q \in (0,1)$.

Suppose not. Then, there exists an $x' > x$ such that $(3*x'-y(x')*q) < (3*x-y(x)*q)$ in SRE for some $q \in (0,1)$. This implies that $3*(x'-x) < q*(y(x')-y(x))$. As actions and beliefs about these actions are the same in any SRE agent B perceives agent A as less kind, when he receives $3*x'$ than when he receives $3*x$. This is true since agent A assigns less expected payoff to agent B, when he chooses x' instead of x:

$$3*x'-y(x')*q < 3*x-y(x)*q,$$

and faces the same reference payoff for x' and x. Hence, $\lambda_B(x') < \lambda_B(x)$. As $3*(x'-x) < q*(y(x')-y(x))$ we know that agent B keeps less for himself when he receives $3*x'$ than when he receives $3*x$:

$$3 * x' - y(x') < 3 * x - y(x),$$

which can be rewritten as

$$3 * (x' - x) < y(x') - y(x).$$

This is true because $q*(y(x') - y(x)) < y(x') - y(x)$. In order that this is, indeed, optimal for agent B the following two (weak) inequalities are necessary:

$$3*x' - y(x') + Y_B * \lambda_B(x') * \big(y(x') - \tfrac{3}{2} * x'\big) \geq 3*x' - y(x) + Y_B * \lambda_B(x') * \big(y(x) - \tfrac{3}{2} * x'\big),$$

because agent B could also send back $y(x) < 3 * x'$ given x' but (weakly) prefers to send back $y(x')$ in this case, and

$$\begin{gathered} 3 * x - y(x) + Y_B * \lambda_B(x) * \big(y(x) - \tfrac{3}{2} * x\big) \geq \\ 3 * x - (y(x') - 3 * (x' - x)) + Y_B * \lambda_B(x) * \big((y(x') - 3 * (x' - x)) - \tfrac{3}{2} * x\big), \end{gathered}$$

because agent B could also send back $(y(x') - 3 * (x' - x)) \leq 3 * x$ given x but (weakly) prefers to send back $y(x)$ in this case. The first (weak) inequality can be rewritten as

$$\lambda_B(x') \geq \tfrac{1}{Y_B},$$

and the second as

$$\lambda_B(x) \leq \tfrac{1}{Y_B}.$$

As $\lambda_B(x') < \lambda_B(x)$ this is a contradiction.

Proposition 1: $\lambda_B(x)$ is (weakly) increasing in x in any SRE, in which agent B chooses a pure strategy, $\forall\ q \in (0, 1)$.

As in any SRE, in which agent B chooses a pure strategy, actions and beliefs about these actions are the same and Lemma 2 holds, agent B beliefs that agent A assigns him more expected material payoff the higher x. Because agent A faces the same reference payoff for any feasible action of his action set, agent B perceives agent A the more kind, the higher x. Therefore, $\lambda_B(x') \geq \lambda_B(x)$ for any feasible pair of $x' > x$.

Proposition 2a: The higher q the (weakly) smaller $y(x, q)$ in any SRE, in which agent B chooses a pure strategy, $\forall\ q \in (0, 1)$ and $x = 20$.

Suppose not. Then, there exists an SRE with $y(20,q)$ for q and an SRE with $y(20,q')$ for $q' > q$ such that $y(20,q) < y(20,q')$. As in any SRE actions and beliefs about these actions are the same agent B beliefs that agent A assigns him more expected payoff when the probability that the game is not stopped is q rather than q':

$$3 * 20 - y(20,q) * q > 3 * 20 - y(20,q') * q'.$$

Does this necessarily imply that agent B perceives agent A as more kind? This depends on the reference payoff for agent A which may vary in q in SRE. The reference payoff for agent A is, again, the mean of the maximal expected material payoff agent A can assign to agent B and the minimal expected material payoff agent A can assign to agent B. Due to Lemma 2 we can simply calculate the reference payoffs in these SREs by dividing the expected payoff agent A assigns to agent B with $x = 20$ by 2. Therefore, the reference payoff is higher in the SRE with q than in that with q'. Nevertheless, agent B perceives agent A as more kind in the SRE with q than in that with q':

$$(3 * 20 - y(20,q) * q) * \tfrac{1}{2} > (3 * 20 - y(20,q') * q') * \tfrac{1}{2}.$$

Hence, $\lambda_B(20,q') < \lambda_B(20,q)$.[40] Agent B, still, gives more in the SRE with q' than in that with q. In order that this is, indeed, optimal for agent B the following two (weak) inequalities are necessary:

$$\begin{gathered} 3 * 20 - y(20,q') + Y_B * \lambda_B(20,q') * \big(y(20,q') - \tfrac{3}{2} * 20\big) \geq \\ 3 * 20 - y(20,q) + Y_B * \lambda_B(20,q') * \big(y(20,q) - \tfrac{3}{2} * 20\big), \end{gathered}$$

because agent B could also send back $y(20,q)$ since he has the same action set as with q but (weakly) prefers to send back $y(20,q')$ in this case, and

$$\begin{gathered} 3 * 20 - y(20,q) + Y_B * \lambda_B(20,q) * \big(y(20,q) - \tfrac{3}{2} * 20\big) \geq \\ 3 * 20 - y(20,q') + Y_B * \lambda_B(20,q) * \big(y(20,q') - \tfrac{3}{2} * 20\big), \end{gathered}$$

because agent B could also send back $y(20,q')$ since he has the same action set as with q' but (weakly) prefers to send back $y(20,q)$ in this case. The first (weak) inequality can be rewritten as

$$\lambda_B(20,q') \geq \tfrac{1}{Y_B},$$

[40] This may not be the case for $x < 20$.

and the second as

$$\lambda_B(20, q) \leq \tfrac{1}{Y_B}.$$

As $\lambda_B(20, q') < \lambda_B(20, q)$ this is a contradiction.

Proposition 2b: $y(20, q)$ in an SRE, in which agent B chooses a pure strategy, equals $y(20, q')$ in an SRE, in which agent B chooses a pure strategy, for $q' > q$, $q', q \in (0, 1)$, if and only if either $y(20, q) = y(20, q') = 60$ or $y(20, q) = y(20, q') = 0$.

Suppose not. Then, there exists an SRE with $y(20, q)$ for q and an SRE with $y(20, q')$ for $q' > q$ such that $60 > y(20, q) = y(20, q') > 0$. As in any SRE actions and beliefs about these actions are the same, agent B beliefs that agent A assigns him more expected payoff when the probability that the game is not stopped is q rather than q':

$$3 * 20 - y(20, q) * q > 3 * 20 - y(20, q') * q'.$$

Does this necessarily imply that agent B perceives agent A as more kind? This depends on the reference payoff for agent A which may vary in q in SRE. The reference payoff for agent A is, again, the mean of the maximal expected material payoff agent A can assign to agent B and the minimal expected material payoff agent A can assign to agent B. Due to Lemma 2, we can simply calculate the reference payoffs in these SREs by dividing the expected payoff agent A assigns to agent B with $x = 20$ by 2. Therefore, the reference payoff is higher in the SRE with q than in that with q'. Nevertheless, agent B perceives agent A as more kind in the SRE with q than in that with q':

$$(3 * 20 - y(20, q) * q) * \tfrac{1}{2} > (3 * 20 - y(20, q') * q') * \tfrac{1}{2}.$$

Hence, $\lambda_B(20, q') < \lambda_B(20, q)$.[41] Agent B, still, gives the same in the SRE with q' than in that with q. In order that this is, indeed, optimal for agent B the following two (weak) inequalities are necessary:

$$3 * 20 - y(20, q') + Y_B * \lambda_B(20, q') * \big(y(20, q') - \tfrac{3}{2} * 20\big) \geq$$
$$3 * 20 - 0 + Y_B * \lambda_B(20, q') * \big(0 - \tfrac{3}{2} * 20\big),$$

[41] This may not be the case for $x < 20$.

because agent B could also send back 0 since he has the same action set as with q but (weakly) prefers to send back $y(20, q')$ in this case, and

$$3 * 20 - y(20, q) + Y_B * \lambda_B(20, q) * \big(y(20, q) - \tfrac{3}{2} * 20\big) \geq$$
$$3 * 20 - 3 * 20 + Y_B * \lambda_B(20, q) * \big(3 * 20 - \tfrac{3}{2} * 20\big),$$

because agent B could also send back $3 * 20$ since he has the same action set as with q' but (weakly) prefers to send back $y(20, q)$ in this case. The first (weak) inequality can be rewritten as

$$\lambda_B(20, q') \geq \tfrac{1}{Y_B},$$

and the second as

$$\lambda_B(20, q) \leq \tfrac{1}{Y_B}.$$

As $\lambda_B(20, q') < \lambda_B(20, q)$ this is a contradiction.

Proposition 3: In any SRE, in which agent B chooses a pure strategy, λ_B is (weakly) larger when $q = 0.5$ than when $q = 0.8$ for $x = 20$.

Suppose not. Then, there exists an SRE with $y(20, 0.5)$ for $q = 0.5$ and an SRE with $y(20, 0.8)$ such that $\lambda_B(20, 0.5) < \lambda_B(20, 0.8)$. Due to Lemma 2 this implies that $y(20, 0.5) > y(20, 0.8)$. In order that this is, indeed, optimal for agent B the following two (weak) inequalities are necessary:

$$3 * 20 - y(20, 0.5) + Y_B * \lambda_B(20, 0.5) * \big(y(20, 0.5) - \tfrac{3}{2} * 20\big) \geq$$
$$3 * 20 - y(20, 0.8) + Y_B * \lambda_B(20, 0.5) * \big(y(20, 0.8) - \tfrac{3}{2} * 20\big),$$

because agent B could also send back $y(20, 0.8)$ since he has the same action set as with $q = 0.5$ but (weakly) prefers to send back $y(20, 0.5)$ in this case, and

$$3 * 20 - y(20, 0.8) + Y_B * \lambda_B(20, 0.8) * \big(y(20, 0.8) - \tfrac{3}{2} * 20\big) \geq$$
$$3 * 20 - y(20, 0.5) + Y_B * \lambda_B(20, 0.8) * \big(y(20, 0.5) - \tfrac{3}{2} * 20\big),$$

because agent B could also send back $y(20, 0.5)$ since he has the same action set as with $q = 0.8$ but (weakly) prefers to send back $y(20, 0.8)$ in this case. The first (weak) inequality can be rewritten as

$$\lambda_B(20, 0.5) \geq \tfrac{1}{Y_B},$$

and the second as

$$\lambda_B(20, 0.8) \leq \frac{1}{Y_B}.$$

As $\lambda_B(20, 0.5) < \lambda_B(20, 0.8)$ this is a contradiction.

Proposition 4: In any SRE, in which agent B chooses a pure strategy, agent A's expected return, $q * y(x, q)$, is (weakly) smaller when $q = 0.5$ than when $q = 0.8$ for $x = 20$.

From Proposition 3 we know that in any SRE, in which agent B chooses a pure strategy, $\lambda_B = \frac{60 - q * y(x,q)}{2}$ is (weakly) larger when $q = 0.5$ than when $q = 0.8$ for $x = 20$. This implies $\frac{60 - 0.5 * y(x,0.5)}{2} > \frac{60 - 0.8 * y(x,0.8)}{2}$ for $x = 20$ which is equivalent to $0.8 * y(x, 0.8) > 0.5 * y(x, 0.5)$ for $x = 20$.

6.2.5 Existence of an SRE

So far, we have developed a couple of statements that hold in any SRE, in which agent B chooses a pure strategy. In the following we show that at least one such SRE exists for each of our treatments.

Lemma 3: $\forall\, x$ and $q \in (0, 1)$ there exists an optimal pure action for agent B such that higher order beliefs about agent B's pure action correspond to this optimal action, i.e. higher order beliefs are correct.

Take an $x < 20$, a $\widetilde{y}(20)$ and the fact that agent A's reference payoff is $\frac{60 - q * \widetilde{y}(20) + 0}{2}$ in any SRE, in which agent B chooses a pure strategy[42]. Agent B's utility function is $U(y(x), \widetilde{y}(x)) = 3 * x - y(x) + Y_B * (3 * x - q * \widetilde{y}(x) - \frac{60 - q * \widetilde{y}(20) + 0}{2}) * \big(y(x) - \frac{3}{2} * x\big)$. U is continuous in $y(x)$ and $\widetilde{y}(x)$, and $U(\cdot, \widetilde{y}(x))$ is quasi-concave. By choosing $y(x) \in G\left(\widetilde{y}(x)\right) = [0, 3 * x]$ agent B can maximize his utility. The correspondence $G\left(\widetilde{y}(x)\right)$ is constant and continuous in $\widetilde{y}(x)$. Furthermore, for any $\widetilde{y}(x)$ $G\left(\widetilde{y}(x)\right)$ is non-empty, compact and convex-valued. Consequently, we can apply Berge's Maximum Theorem and conclude that for any $\widetilde{y}(x) \in [0, 3 * x]$ there exists a $y(x) \in [0, 3 * x]$ that maximizes $U(y(x), \widetilde{y}(x))$ and the correspondence $Y^*(\widetilde{y}(x)) : [0, 3 * x] \rightarrow [0, 3 * x]$ that maps $\widetilde{y}(x) \in [0, 3 * x]$ into the set of $y(x) \in [0, 3 * x]$ that maximize $U(y(x), \widetilde{y}(x))$ is non-empty, compact-valued, upper-hemicontinuous and convex-valued. It remains to show that $Y^*(\widetilde{y}(x))$ has a fixed point $\widetilde{y}(x) \in Y^*(\widetilde{y}(x))$,

[42] This is due to Lemma 2.

i.e. higher order beliefs are correct. We apply Kakutani's Fixed Point Theorem and conclude that at least one fixed point exists.

For $x = 20$ and the fact that agent A's reference payoff is $\frac{60-q*\widetilde{y}(20)+0}{2}$ in any SRE, in which agent B chooses a pure strategy[43], agent B's utility function is $U(y(20), \widetilde{y}(20)) = 3 * 20 - y(20) + Y_B * \lambda_B * \frac{1}{2} * (60 - q * \widetilde{y}(20)) * f(y(20) - 30)$. Again, U is continuous in $y(20)$ and $\widetilde{y}(20)$, $U(\cdot, \widetilde{y}(20))$ is quasi-concave, and $[0, 60]$ is continuous in $\widetilde{y}(20)$, non-empty, compact and convex-valued. As above we can conclude that for any $\widetilde{y}(20) \in [0, 60]$ there exists a $y(20) \in [0, 60]$ that maximizes $U(y(20), \widetilde{y}(20))$ and that there exist at least one higher order belief $\widetilde{y}(20)$ that is correct.

Proposition 5: For any $q \in (0, 1)$ there exists an SRE, in which agent B chooses a pure strategy.

Due to Lemma 3 it remains to show that given agent B's pure optimal strategy agent A has an optimal (possibly randomized) strategy that is correctly expected. Take any pure strategy of agent B. Let agent A's utility function be $U(a, \widetilde{a}) = \pi_A + Y_A * \kappa_A * \lambda_A$, with $a \in X$ as agent A's (possibly randomized) action, $\widetilde{a} \in X$ as agent A's second order belief on a, and X as agent A's set of possibly randomized actions. $U(a, \widetilde{a})$ is continuous, $U(\cdot, \widetilde{a})$ is quasi-concave, and X is continuous in $\widetilde{a}$, non-empty, compact and convex-valued. Hence, we can apply Berge's Maximum Theorem and conclude that for any $\widetilde{a}$ there exists a set of actions $X^*(\widetilde{a})$ out of which each action is part of the set X and maximizes agent A's utility given $\widetilde{a}$. Furthermore, $X^*(\widetilde{a}) : X \to X$ is a non-empty, compact, convex-valued upper-hemicontinuous correspondence. Consequently, we can apply Kakutani's Fixed Point Theorem and conclude that $X^*(\widetilde{a})$ has at least one fixed point.

6.3 Behavioral predictions of the modified version of the model by Dufwenberg and Kirchsteiger (2004) with central elements of the model by Falk and Fischbacher (2006)

6.3.1 The model

We consider the same utility function of individual i as from the modified version of the model by Dufwenberg and Kirchsteiger (2004), but define κ_i and λ_i differently.

[43] This is due to Lemma 2.

The interpretation of these terms, though, remains the same. The reference payoffs in these two terms change: We define the reference payoff for κ_i as individual i's expected material payoff, π_i, and the reference payoff for λ_i as the individual i's expectation of individual j's expected payoff. The updating of beliefs and the SRE are the same as in the modified version of the model by Dufwenberg and Kirchsteiger (2004).

6.3.2 Agent B's utility function when he is asked to decide

Consider agent A has chosen on x and the lottery has determined to continue the game. Agent B, then, decides on $y(x) \in [0, 3 * x]$ and beliefs that agent A has chosen $x > 0$ (with probability 1), that nature has chosen to continue the game (with probability 1) and that agent A beliefs that agent B returns $\widetilde{y}(0) = 0$, $\widetilde{y}(5) \in [0, 15]$, $\widetilde{y}(10) \in [0, 30]$, $\widetilde{y}(15) \in [0, 45]$, $\widetilde{y}(20) \in [0, 60]$, where $\widetilde{y}(x)$ represents agent B's second order belief for x.

Then, π_B, agent B's expected own material payoff, is equal to $3*x-y(x)$ and κ_B, agent B's perception of the kindness of his own strategy, is equal to $(20 - x + y(x)) - (3 * x - y(x))$. λ_B, agent B's perception of the kindness of agent A's strategy, is equal to $(3 * x - q * \widetilde{y}(x)) - (20 - x + q * \widetilde{y}(x))$.

6.3.3 Equilibrium predictions

In this subsection we derive some statements that hold in any SRE, in which agent B chooses a pure strategy.

Lemma 1a': $y(x)$ is (weakly) increasing in x in every SRE, in which agent B chooses a pure strategy, $\forall\ q \in (0, 1)$.

Suppose not. Then, there exists an $x' > x$ such that $y(x') < y(x)$ in SRE.. Because in any SRE actions and beliefs about these actions are the same, agent B perceives agent A as more kind, when he receives $3 * x'$ than when he receives $3 * x$. This is true since agent A assigns more expected payoff to agent B, when he chooses x' instead of x:

$$3 * x' - y(x') * q > 3 * x - y(x) * q,$$

and faces a smaller reference payoff for x' than for x:

$$20 - x' + y(x') * q < 20 - x + y(x) * q.$$

Hence, $\lambda_B(x') > \lambda_B(x)$. Nevertheless, agent B returns less when he receives $3*x'$ than when he receives $3*x$. In order that this is, indeed, optimal for agent B the following two (weak) inequalities are necessary:

$$3*x' - y(x') + Y_B * \lambda_B(x') * (20 - 4*x' + 2*y(x')) \geq$$
$$3*x' - y(x) + Y_B * \lambda_B(x') * (20 - 4*x' + 2*y(x)),$$

because agent B could also send back $y(x) < 3*x'$ given x' but (weakly) prefers to send back $y(x')$ in this case, and

$$3*x - y(x) + Y_B * \lambda_B(x) * (20 - 4*x + 2*y(x)) \geq$$
$$3*x - y(x') + Y_B * \lambda_B(x) * (20 - 4*x + 2*y(x')),$$

because agent B could also send back $y(x') < y(x)$ given x but (weakly) prefers to send back $y(x)$ in this case. The first (weak) inequality can be rewritten as

$$\tfrac{1}{2*Y_B} \geq \lambda_B(x'),$$

and the second as

$$\tfrac{1}{2*Y_B} \leq \lambda_B(x).$$

As $\lambda_B(x') > \lambda_B(x)$ this is a contradiction.

Lemma 1b': $\widetilde{y}(x)$ is (weakly) increasing in x in any SRE, in which agent B chooses a pure strategy, $\forall\ q \in (0,1)$.

As in any SRE, in which agent B chooses a pure strategy, actions and beliefs about these actions are the same, $y(x) = \widetilde{y}(x)\ \forall\ x$. Due to Lemma 1a' $y(x)$ is increasing in x in any SRE and therefore, also $\widetilde{y}(x)$.

Lemma 2': $(3*x - y(x)*q)$ is (weakly) increasing in x in every SRE, in which agent B chooses a pure strategy, $\forall\ q \in (0,1)$.

Suppose not. Then, there exists an $x' > x$ such that $(3*x' - y(x')*q) < (3*x - y(x)*q)$ in SRE. This implies that $3*(x'-x) < q*(y(x') - y(x))$. As actions and beliefs about these actions are the same in any SRE agent B perceives agent A as less kind when he receives $3*x'$ than when he receives $3*x$. This is true since agent A assigns less expected payoff to agent B when he chooses x' instead of x:

$$3 * x' - y(x') * q < 3 * x - y(x) * q,$$

and faces a larger reference payoff for x' than for x:

$$20 - x' + y(x') * q > 20 - x + y(x) * q,$$

which can be rewritten as

$$q * (y(x') - y(x)) > x' - x.$$

The above inequality holds since $q*(y(x') - y(x)) > 3*(x' - x)$. Hence, $\lambda_B(x') < \lambda_B(x)$. As $3 * (x' - x) < q * (y(x') - y(x))$ we also know that agent B keeps less when he receives $3 * x'$ and more when he receives $3 * x$:

$$3 * x' - y(x') < 3 * x - y(x),$$

which can be rewritten as

$$3 * (x' - x) < y(x') - y(x).$$

This is true because $q*(y(x') - y(x)) < y(x') - y(x)$. In order that this is, indeed, optimal for agent B the following two (weak) inequalities are necessary:

$$3 * x' - y(x') + Y_B * \lambda_B(x') * (20 - 4 * x' + 2 * y(x')) \geq$$
$$3 * x' - y(x) + Y_B * \lambda_B(x') * (20 - 4 * x' + 2 * y(x)),$$

because agent B could also send back $y(x) < 3 * x'$ given x' but (weakly) prefers to send back $y(x')$ in this case, and

$$3 * x - y(x) + Y_B * \lambda_B(x) * (20 - 4 * x + 2 * y(x)) \geq$$
$$3 * x - (y(x') - 3 * (x' - x)) + Y_B * \lambda_B(x) * (20 - 4 * x + 2 * (y(x') - 3 * (x' - x))),$$

because agent B could also send back $(y(x') - 3 * (x' - x)) \leq 3 * x$ given x but (weakly) prefers to send back $y(x)$ in this case. The first (weak) inequality can be rewritten as

$$\lambda_B(x') \geq \frac{1}{2 * Y_B},$$

and the second as

$$\lambda_B(x) \leq \frac{1}{2 * Y_B}.$$

As $\lambda_B(x') < \lambda_B(x)$ this is a contradiction.

Proposition 1': $\lambda_B(x)$ is (weakly) increasing in x in every SRE, in which agent B chooses a pure strategy, $\forall\ q \in \{0.5, 0.8\}$.

Suppose not. Then, there exists an $x' > x$ such that $\lambda_B(x') < \lambda_B(x)$ in SRE. Due to Lemma 1a', $y(x') > y(x)$ in any SRE, in which agent B chooses a pure strategy. Furthermore, we know that when $q \in \{0.5, 0.8\}$ $y(x) < 3 * x$ in SRE.[44] Hence, agent B has the opportunity to return more than $y(x)$ given x. In order that the behavior as described above is, indeed, optimal for agent B the following two (weak) inequalities are necessary:

$$3 * x' - y(x') + Y_B * \lambda_B(x') * (20 - 4 * x' + 2 * y(x')) \geq$$
$$3 * x' - y(x) + Y_B * \lambda_B(x') * (20 - 4 * x' + 2 * y(x)),$$

because agent B could also send back $y(x) < 3 * x'$ given x' but (weakly) prefers to send back $y(x')$ in this case, and

$$3 * x - y(x) + Y_B * \lambda_B(x) * (20 - 4 * x + 2 * y(x)) \geq$$
$$3 * x - 3 * x + Y_B * \lambda_B(x) * (20 - 4 * x + 2 * 3 * x),$$

because agent B could also send back $3 * x > y(x)$ given x but (weakly) prefers to send back $y(x)$ in this case. The first (weak) inequality can be rewritten as

$$\lambda_B(x') \geq \tfrac{1}{2 * Y_B}$$

and the second as

$$\lambda_B(x) \leq \tfrac{1}{2 * Y_B}$$

As $\lambda_B(x') < \lambda_B(x)$ this is a contradiction.

Proposition 2a': The higher q the (weakly) smaller $y(x, q)$ in any SRE, in which agent B chooses a pure strategy, $\forall\ q \in (0, 1)$ and $x > 0$.

Suppose not. Then, there exists an SRE, in which agent B chooses a pure strategy**,** with $y(x, q)$ for q and $x > 0$ and an SRE, in which agent B chooses a pure strategy**,** with $y(x, q')$ for $q' > q$ and $x > 0$ such that $y(x, q) < y(x, q')$. As in any SRE actions and beliefs about these actions are the same, agent B beliefs that agent A assigns him more expected payoff when the probability that the game is not stopped is q rather than q':

[44] If not and $y(x) = \widetilde{y}(x) = 3 * x$, $\lambda_B(x) = 4 * x - 20 - 2 * \widetilde{y}(x) * q$, which is smaller than 0 for $q \in \{0.5, 0.8\}$, and agent B preferred to return nothing.

$$3 * x - y(x, q) * q > 3 * x - y(x, q') * q'.$$

Furthermore, the reference payoff for agent A is smaller when the probability that the game is not stopped is q rather than q':

$$20 - x + y(x, q) * q < 20 - x + y(x, q') * q'.$$

Hence, $\lambda_B(x, q') < \lambda_B(x, q)$. Again, this is true $\forall x > 0$. Agent B, still, gives more in the SRE with q' than in that with q. In order that this is, indeed, optimal for agent B the following two (weak) inequalities are necessary:

$$\begin{gathered} 3 * x - y(x, q') + Y_B * \lambda_B(x, q') * (20 - 4 * x + 2 * y(x, q')) \geq \\ 3 * x - y(x, q) + Y_B * \lambda_B(x, q') * (20 - 4 * x + 2 * y(x, q)), \end{gathered}$$

because agent B could also send back $y(x, q)$ since he has the same action set as with q but (weakly) prefers to send back $y(x, q')$ in this case, and

$$\begin{gathered} 3 * x - y(x, q) + Y_B * \lambda_B(x, q) * (20 - 4 * x + 2 * y(x, q)) \geq \\ 3 * x - y(x, q') + Y_B * \lambda_B(x, q) * (20 - 4 * x + 2 * y(x, q')), \end{gathered}$$

because agent B could also send back $y(20, q')$ since he has the same action set as with q' but (weakly) prefers to send back $y(20, q)$ in this case. The first (weak) inequality can be rewritten as

$$\lambda_B(x, q') \geq \tfrac{1}{2 * Y_B},$$

and the second as

$$\lambda_B(x, q) \leq \tfrac{1}{2 * Y_B}.$$

As $\lambda_B(x, q') < \lambda_B(x, q)$ this is a contradiction.

Proposition 2b': For $x > 0$ $y(x, q)$ in an SRE, in which agent B chooses a pure strategy, equals $y(x, q')$ in an SRE, in which agent B chooses a pure strategy, for $q' > q$, $q', q \in (0, 1)$, if and only if either $y(x, q) = y(x, q') = 3 * x$ or $y(x, q) = y(x, q') = 0$.

Suppose not. Then, there exists an SRE with $y(x, q)$ for q and an SRE with $y(x, q')$ for $q' > q$ such that $3 * x > y(x, q) = y(x, q') > 0$. As in any SRE actions and beliefs about these actions are the same, agent B beliefs that agent A assigns him more expected payoff when the probability that the game is not stopped is q rather than q':

$$3 * x - y(x, q) * q > 3 * x - y(x, q') * q'.$$

Furthermore, the reference payoff for agent A is smaller when the probability that the game is not stopped is q rather than q':

$$20 - x + y(x, q) * q < 20 - x + y(x, q') * q'.$$

Hence, $\lambda_B(x, q') < \lambda_B(x, q)$. Again, this is true $\forall\ x > 0$. Agent B, still, gives the same in the SRE with q' than in that with q. In order that this is, indeed, optimal for agent B the following two equalities are necessary:

$$\begin{gathered} 3 * x - y(x, q') + Y_B * \lambda_B(x, q') * (20 - 4 * x + 2 * y(x, q')) \geq \\ 3 * x - 0 + Y_B * \lambda_B(x, q') * (20 - 4 * x + 2 * 0), \end{gathered}$$

because agent B could also send back 0 since he has the same action set as with q but (weakly) prefers to send back $y(x, q')$ in this case, and

$$\begin{gathered} 3 * x - y(x, q) + Y_B * \lambda_B(x, q) * (20 - 4 * x + 2 * y(x, q)) \geq \\ 3 * x - 3 * x + Y_B * \lambda_B(x, q) * (20 - 4 * x + 2 * 3 * x), \end{gathered}$$

because agent B could also send back $3 * x$ since he has the same action set as with q' but (weakly) prefers to send back $y(x, q)$ in this case. The first (weak) inequality can be rewritten as

$$\lambda_B(x, q') \geq \tfrac{1}{2 * Y_B},$$

and the second as

$$\lambda_B(x, q) \leq \tfrac{1}{2 * Y_B}.$$

As $\lambda_B(x, q') < \lambda_B(x, q)$ this is a contradiction.

Proposition 3': In any SRE, in which agent B chooses a pure strategy, λ_B is (weakly) larger when $q = 0.5$ than when $q = 0.8$ for $x > 0$.

Suppose not. Then, there exists an SRE with $y(x, 0.5)$ for $q = 0.5$ and an SRE with $y(x, 0.8)$ for an $x > 0$ such that $\lambda_A(x, 0.5) < \lambda_A(x, 0.8)$. This implies that $y(x, 0.5) > y(x, 0.8)$. In order that this is, indeed, optimal for agent B the following two (weak) inequalities are necessary:

$$\begin{gathered} 3 * x - y(x, 0.5) + Y_B * \lambda_B(x, 0.5) * (20 - 4 * x + 2 * y(x, 0.5)) \geq \\ 3 * x - y(x, 0.8) + Y_B * \lambda_B(x, 0.5) * (20 - 4 * x + 2 * y(x, 0.8)), \end{gathered}$$

because agent B could also send back $y(x, 0.8)$ since he has the same action set as with $q = 0.8$ but (weakly) prefers to send back $y(x, 0.5)$ in this case, and

$$3 * x - y(x, 0.8) + Y_B * \lambda_B(x, 0.8) * (20 - 4 * x + 2 * y(x, 0.8)) \geq$$
$$3 * x - y(x, 0.5) + Y_B * \lambda_B(x, 0.8) * (20 - 4 * x + 2 * y(x, 0.5)),$$

because agent B could also send back $y(x, 0.5)$ since he has the same action set as with $q = 0.5$ but (weakly) prefers to send back $y(x, 0.8)$ in this case. The first (weak) inequality can be rewritten as

$$\lambda_B(x, 0.5) \geq \frac{1}{2 * Y_B},$$

and the second as

$$\lambda_B(x, 0.8) \leq \frac{1}{2 * Y_B}.$$

As $\lambda_B(x, 0.5) < \lambda_B(x, 0.8)$ this is a contradiction.

Proposition 4': In any SRE, in which agent B chooses a pure strategy, agent A's expected return, $q * y(x, q)$, is (weakly) smaller when $q = 0.5$ than when $q = 0.8$ for $x > 0$.

From Proposition 3' we know that in any SRE, in which agent B chooses a pure strategy, $\lambda_B = 4 * x - 2 * q * y(x, q) - 20$ is (weakly) larger when $q = 0.5$ than when $q = 0.8$ for $x > 0$. This implies $4 * x - 2 * 0.5 * y(x, 0.5) - 20 > 4 * x - 2 * 0.8 * y(x, 0.8) - 20$ for $x > 0$ which is equivalent to $0.8 * y(x, 0.8) > 0.5 * y(x, 0.5)$ for $x > 0$.

6.3.4 Existence of an SRE

So far, we have developed a couple of statements that hold in any SRE, in which agent B chooses a pure strategy. In the following we show that at least one such SRE exists for each of our treatments.

Lemma 3': $\forall$ x and $q \in (0, 1)$ there exists an optimal pure action for agent B such that higher order beliefs about agent B's pure action correspond to this optimal action, i.e. higher order beliefs are correct.

Take an x and the fact that agent A's reference payoff is $20 - x + q * \widetilde{y}(x)$. Agent B's utility function is $U(y(x), \widetilde{y}(x)) = 3 * x - y(x) + Y_B * (3 * x - q * \widetilde{y}(x) - 20 + x - q * \widetilde{y}(x)) * (20 - 4 * x + 2 * q * y(x))$. U is continuous in $y(x)$ and $\widetilde{y}(x)$, and

$U(\cdot, \widetilde{y}(x))$ is quasi-concave. By choosing $y(x) \in G(\widetilde{y}(x)) = [0, 3 * x]$ agent B can maximize his utility. The correspondence $G(\widetilde{y}(x))$ is constant and continuous in $\widetilde{y}(x)$. Furthermore, for any $\widetilde{y}(x)$ $G(\widetilde{y}(x))$ is non-empty, compact and convex-valued. Consequently, we can apply Berge's Maximum Theorem and conclude that for any $\widetilde{y}(x) \in [0, 3 * x]$ there exists a set of $y(x) \in [0, 3 * x]$ that maximize $U(y(x), \widetilde{y}(x))$ and the correspondence $Y^*(\widetilde{y}(x)) : [0, 3 * x] \to [0, 3 * x]$ that maps $\widetilde{y}(x) \in [0, 3 * x]$ into the $y(x) \in [0, 3 * x]$ that maximizes $U(y(x), \widetilde{y}(x))$ is non-empty, compact-valued, upper-hemicontinuous and convex-valued. It remains to show that $Y^*(\widetilde{y}(x))$ has a fixed point $\widetilde{y}(x) \in Y^*(\widetilde{y}(x))$, i.e. higher order beliefs are correct. We apply Kakutani's Fixed Point Theorem and conclude that at least one fixed point exists.

Proposition 5': For any $q \in (0, 1)$ there exists an SRE, in which agent B chooses a pure strategy.

Due to Lemma 3' it remains to show that given agent B's pure optimal strategy agent A has an optimal (possibly randomized) strategy that is correctly expected. Take any pure strategy of agent B. Let agent A's utility function be $U(a, \widetilde{a}) = \pi_A + Y_A * \kappa_A * \lambda_A$, with $a \in X$ as agent A's (possibly randomized) action, $\widetilde{a} \in X$ as agent A's higher order belief on a, and X as agent A's set of possibly randomized actions. $U(a, \widetilde{a})$ is continuous, $U(\cdot, \widetilde{a})$ is quasi-concave, and X is continuous in $\widetilde{a}$, non-empty, compact and convex-valued. Hence, we can apply Berge's Maximum Theorem and conclude that for any $\widetilde{a}$ there exists a set of actions $X^*(\widetilde{a})$ out of which each action is part of the set X and maximizes agent A's utility given $\widetilde{a}$. Furthermore, $X^*(\widetilde{a}) : X \to X$ is a non-empty, compact, convex-valued upper-hemicontinuous correspondence. Consequently, we can apply Kakutani's Fixed Point Theorem and conclude that $X^*(\widetilde{a})$ has at least one fixed point.

References

[1] Anderson, L. R., Holt, C. A., 1997. Information cascades in the laboratory. American Economic Review 87, 847-862.

[2] Andreoni, J., Miller, J., 2002. Giving according to GARP: An experimental test of the rationality of altruism. Econometrica 70, 737-753.

[3] Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games and Economic Behavior 10, 122-142.

[4] Blount, S., 1995. When social outcomes aren't fair: The effect of causal attributions on preferences. Organizational Behavior and Human Decision Processes 63, 131-144.

[5] Bolton, G. E., Ockenfels, A., 2000. A theory of equity, reciprocity and competition. American Economic Review 90, 166-193.

[6] Camerer, C. F., 2003. Behavioral game theory. Experiments in strategic interaction. Princeton: Princeton University Press.

[7] Charness, G., 2004. Attribution and reciprocity in an experimental labor market. Journal of Labor Economics 22, 665-688.

[8] Cox, J. C., 2004. How to identify trust and reciprocity. Games and Economic Behavior 46, 260-281.

[9] Dufwenberg, M., Kirchsteiger, M., 2004. A theory of sequential reciprocity. Games and Economic Behavior 47, 268-298.

[10] Goeree, J., Palfrey, T. R., Rogers, B. W., McKelvey, R. D., 2007. Self-Correcting information cascades. Review of Economic Studies 74, 733-762.

[11] Falk, A., Fehr, E., Fischbacher, U., 2003. On the nature of fair behavior. Economic Inquiry 41, 20-26.

[12] Falk, A., Fehr, E., Fischbacher, U., 2008. Testing theories of fairness - Intentions matter. Games and Economic Behavior 62, 287-303.

[13] Falk, A., Fischbacher U., 2006. A theory of reciprocity. Games and Economic Behavior 54, 293-315.

[14] Fehr, E., Schmidt, K. M., 1999. A theory of fairness, competition, and cooperation. Quarterly Journal of Economics 114, 817-868.

[15] Fischbacher, U., 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. Experimental Economics 10, 171-178.

[16] Forsythe, R., Horowitz, J. L., Savin, N. E., Sefton, M., 1994. Fairness in simple bargaining experiments. Games and Economic Behavior 6, 347-369.

[17] Hung, A. A., Plott, C. R., 2001. Information cascades: Replication and an extension to majority rule and conformity-rewarding. American Economic Review 91, 1508-1520.

[18] Kariv, S., 2005. Overconfidence and informational cascades. Working Paper.

[19] Nöth, M., Weber, M., 2003. Information aggregation with random ordering: Cascades and overconfidence. The Economic Journal 113, 166-189.

[20] Offerman, T., 2002. Hurting hurts more than helping helps. European Economic Review 46, 1423-1437.

[21] Rabin, M., 1993. Incorporating fairness into game game theory and economics. American Economic Review 83, 1281-1302.

[22] Stanca, L., Bruni, L., Corazzini, L., 2009. Testing theories of reciprocity: Does motivation matter? Journal of Economic Behavior and Organization. forthcoming.

[23] Trautmann, S. T., 2009. A tractable model of process fairness under risk. Working Paper.