GOVERNANCE AND THE EFFICIENCY OF ECONOMIC SYSTEMS

GESY

Discussion Paper No. 302

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

Christina Strassmair*

* University of Munichy

October 2009

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

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

Christina Strassmair
University of Munich†

October 7, 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 expected rewards rather than by pure altruism. The question we address in this experimental study is how B reacts to A's intentions. We vary the probability that the second mover in a trust game can reciprocate and analyze effects on second mover behavior. Our results suggest that expected rewards do not spoil the perceived kindness of an action and the action's rewards.

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

JEL classification: C91, D03, D64

*I thank Klaus Abbink, Georg Gebhardt, Martin Kocher, Sandra Ludwig, Daniele Nosenzo, 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, the Theory Workshop at University of Munich, the Annual Congress of the European Economic Association 2009 in Barcelona, and the Economic Science Association – European Meeting 2009 in Innsbruck 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, gratefully acknowledged. I kindly thank the Center for Experimental Economics of the University of Innsbruck for providing laboratory resources, CREED of the University of Amsterdam for the hospitality, and the SFB TR/15.

†*Affiliation:* University of Munich, Seminar for Economic Theory, Ludwigstraße 28 (Rgb.), 80539 Munich, Germany. *Tel.:* +49 89 2180 2926, *Fax:* +49 89 2180 3510. *E-mail:* christina.strassmair@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 study 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. Cox (2004), for example, documents strategic behavior of first movers in a trust game. Other examples are companies that give Christmas gifts to their business partners in order to improve their 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 perceived 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 that the second mover can reciprocate and analyze effects on second mover behavior. Our results suggest that expected future rewards do not spoil the perceived kindness of the first mover's action and the rewards given by the second mover.

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. Agent B receives the tripled amount of agent A's transfer. Then, a lottery determines whether agent B can decide on his return transfer to agent A or not. In the latter case nothing is returned to agent A. We conduct two treatments of this modified trust game that differ in the probability that agent B can decide on his return transfer: In treatment T-HIGH this probability is 80 % and in treatment T-LOW it is 50 %. In both treatments agent A behaves in a way that seems altruistic when he transfers a strictly positive amount to agent B. Our treatment variation, however, changes the possibility for agent B to make a return transfer to agent A and, thereby, varies the chance for agent A to receive a return. Our models of intention-based reciprocity that build on Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006) predict that agent A *ex ante* expects smaller future rewards for a given transfer

in T-LOW than in T-HIGH and, therefore, that agent B returns more in T-LOW when he is asked to decide. In equilibrium the difference in the probability that agent B can decide on his return transfer dominates the difference in agent B's return transfer.

Our results suggest that expected future returns do not spoil the perceived kindness of an action and the action's rewards. Agent B's return transfer (for a given transfer of agent A) does not differ across treatments. This is not because agent B does not care about agent A's action at all. Actually, we observe a lot of agents B who return strictly positive amounts and, in addition, agent B's average return transfer increases in agent A's transfer. This suggests that individuals reward actions that seem altruistic, irrespective of the actor's expectation of future rewards. We conclude that individuals in our setting condition their behavior on outcomes rather than on intentions.

In a next step we formulate possible explanations for our findings and evaluate them with data from our questionnaire that participants filled out after they made their decisions. First, our regressions of agent B's return transfer on agent B's elicited second order belief give no indication that actual (possibly incorrect) expected future returns spoil the kindness of an action and the action's rewards. Hence, incorrect higher order beliefs do not seem to be a plausible explanation for our findings. Second, we analyze treatment differences in agent B's perception of agent A's action and in agent B's emotions. The former seems hardly affected by the treatment, while there is some evidence that appreciation (anger and contempt) is more (less) strongly experienced 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 modeled in a number of theoretical papers. Rabin (1993), Dufwenberg and Kirchsteiger (2004), and Falk and Fischbacher (2006) introduce theoretical models of intention-based reciprocity. In contrast to models of social preferences that are based on outcomes only (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), they (also) take into account that intentions affect the perception of others' actions and, thereby, behavior.

Various experimental papers have focused on the empirical relevance of intentions. A couple of them (e.g. Blount, 1995; Offerman, 2002; McCabe et al., 2003; Charness, 2004; Cox, 2004; Falk et al., 2008) study the effect of intentionality, i.e. whether the second mover's reaction to the first mover's action is different when the first mover's action is chosen by the first mover himself (out of a non-singleton action

set) than when it is exogenously determined by an experimenter or a lottery. These studies typically find that the second mover returns the favor or the disfavor of the first mover's action in a more pronounced way when it was intentionally chosen by the first mover himself. Hence, intentionality seems to matter. These studies, however, do not provide evidence for the effect of *different* intentions behind the *same intentional* action.

Charness and Levine (2007) go further in this direction. They present experimental results of a gift exchange game in which the principal determines an initial wage that is subsequently hit by a random shock. Agents work less for the *same* final wage when it was brought about by a lousy initial wage of the principal and a positive shock than by a generous initial wage and a negative shock. While this study compares two different intentional actions that lead to the same outcome, Bolton et al. (1998) and Falk et al. (2003) compare the reaction to the same intentional action when different (non-singleton) action sets are available for the first mover. Typically, the same action is either the most or the least generous action of the first mover's action set. While Bolton et al. (1998) do not observe any significant effects in their experimental setting, Falk et al. (2003) find in an experimental ultimatum game that responders reject the same offer less often when it is the most generous offer of the first mover's action set than when it is the least generous. Hence, there is evidence that the relative position of an action in the first mover's choice set seems to matter. In our study, in contrast, we focus on gifts, i.e. on intentional actions that always seem to be altruistic or generous. In all of our treatments the first mover's action set is the same and so is the ranking of the actions' generosity. We only vary the second mover's possibility to reciprocate and, thereby, can study different intentions behind the same gift. In particular, we ask whether expected rewards spoil the kindness of a gift.

Stanca et al. (forthcoming) analyze in their experimental study whether the second mover's reaction differs when the first mover's action is extrinsically motivated rather than intrinsically. They compare the second mover's reaction in a standard trust game with the corresponding reaction in a trust game in which first movers are not informed that second movers can react to their transfer until they made their decision.[1] 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

[1]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 and there are no surprises with respect to the underlying game.

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 can directly test models of intention-based reciprocity that predict that the second mover returns more *for a given transfer* in T-LOW than in T-HIGH.[2]

This 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 a 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, however, the game continues and agent B can transfer an amount $y(x) \in [0, 3 * x]$ to agent A. In the latter case 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, $3 * x - y(x)$. 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. If $q = 1$, the game resembles a trust game. If $q = 0$, the game boils down to a dictator game. 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 a 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.

Since we are interested in the effect of an actor's expectation to receive future rewards on the perceived kindness of his action and on the action's rewards, we vary q, the probability that agent B can make a return transfer (given $x > 0$), across treatments and keep everything else constant. Table 1 presents our treatments.

[2] This hypothesis 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 modified trust game

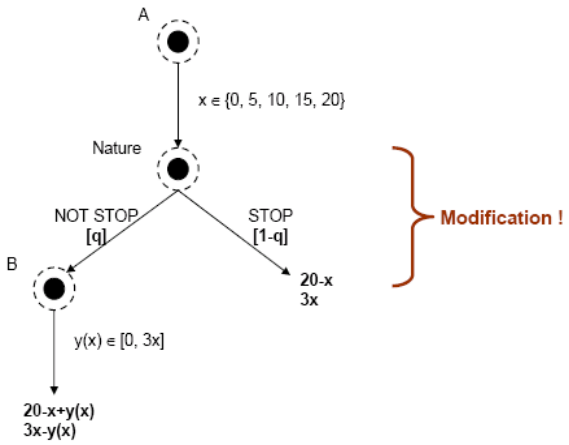


Table 1: Treatments of the modified trust game

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 equal (or close) to 0 and 1 since we would like to compare settings all with non-degenerate lotteries. Furthermore, we are restrained to take higher values of q since agents B are only asked to decide on $y(x)$ when the game continues. If instead q was small, we expected only very few observations of $y(x)$ for a given number of participants.[3]

In each experimental session one treatment of the modified trust game is conducted. The implemented treatment 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. Given agents B are asked to

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

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.[4] After 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 sex, age, and subject of studies.

Our experimental sessions were run at the Center for Experimental Economics of the University of Innsbruck, Austria, in April 2008. 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 and lasted about an hour.[5] Individuals earned on average 10.34 € (at the time of the experiment 1 € ≈ 1.57 USD) including a show-up fee of 5 €. The maximum payoff was 23 € and the minimum 5 €.

3 Behavioral predictions and hypotheses

An actor's expectation to receive future rewards may spoil the kindness of his gift (i.e. his costly action that benefits others) and the gift's rewards because future rewards can partially cover the actor's initial costs and reduce the others' net benefit. In the presented modified trust game agent A behaves in a way that seems altruistic when he transfers a strictly positive amount to agent B: Agent A's transfer is costly – it is deducted from his initial endowment – and benefits agent B. This is true for both treatments. Our treatment variation, however, changes the possibility for agent B to make a return transfer and, thereby, varies the chance for agent A to receive a future return for his transfer. Hence, agent A's expected returns (for a given transfer) are smaller in T-LOW, given agent A has the *same* belief about agent B's reaction in both treatments. Consequently, less of agent A's initial costs are covered in expectation in T-LOW, more expected payoff is assigned to agent B in T-LOW, and, therefore, agent B may perceive agent A as kinder in T-LOW and, in fact, return more when he is asked to decide. If, however, agent A's belief about agent B's reaction is correct and agent B's belief about agent A's belief is correct,

[4] 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. (forthcoming) argue that the strategy method applied in their trust games does not significantly affect decisions.

[5] Translated instructions and a more detailed description of the session's procedure are provided in the appendix.

then agent A expects agent B to transfer more in T-LOW when agent B is asked to decide. Nevertheless, agent A ex ante faces less expected future returns in T-LOW since the difference in q compensates the difference in agent B's reaction. Why is this the case? Suppose the opposite: Agent A ex ante expected higher future returns in T-LOW. Then, agent B perceived agent A as less kind in T-LOW and transferred less in T-LOW. Consequently, agent A's expectation about his future returns were incorrect.

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

3.1 Behavioral predictions of agent B

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 increases in m.

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

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

A model of social preferences based on outcomes

Models of social preferences based on outcomes (e.g. Fehr and Schmidt, 1999; 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 seeking.

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

Since agent B's decision is affected by outcomes only (and not 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.[6]

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

Models of intention-based reciprocity

Models of intention-based reciprocity (e.g. 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 and on the other individuals' intentions. An individual's intentions shape the kindness of his action. How kindness is defined exactly and how intentions concretely enter the utility function varies across the different models. Typically, the kinder an individual perceives another individual's action, the kinder the individual treats this other individual. We set up a model that is based on Dufwenberg and Kirchsteiger (2004). In one specification we use a definition for kindness that is similar to the one by Dufwenberg and Kirchsteiger (2004), DK-specification, in another specification it is similar to the one by Falk and Fischbacher (2006), FF-specification.[7]

Given the assumptions of the model, $y(x)$ is (weakly) higher in T-LOW than in T-HIGH $\forall\ x \in A$ in any sequential reciprocity equilibrium (SRE) in which agent B chooses a pure strategy.[8] In the DK-specification $A = \{20\}$ and in the FF-specification $A = \{0, 5, 10, 15, 20\}$.[9]

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.

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.

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 $x > 0$.

[7] In the appendix we introduce these models and derive their predictions.

[8] 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.

[9] The reason for the possibly different prediction is the difference in the reference action that is used for the definition of kindness.

Hypothesis 1 is supported by the self-interest model: Actions that seem altruistic are never rewarded. Models of social preferences based on outcomes support Hypothesis 2. Similar to the self-interest model, there are no treatment effects *with respect to agent B's behavior*. In contrast to the self-interest model, however, agent B returns a weakly positive amount to agent A: Actions that seem altruistic are rewarded, irrespective of the actor's intentions. Our models of intention-based reciprocity take into consideration the effect of intentions. They predict that the expectation of future returns spoils the kindness of an action that seems altruistic and the action's rewards. Hence, they support Hypothesis 3.[10]

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 expectation of future rewards spoils the kindness of an action that seems altruistic and the action's rewards. Finally, we examine our questionnaire data to evaluate possible explanations for our findings.

4.1 Summary statistics

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

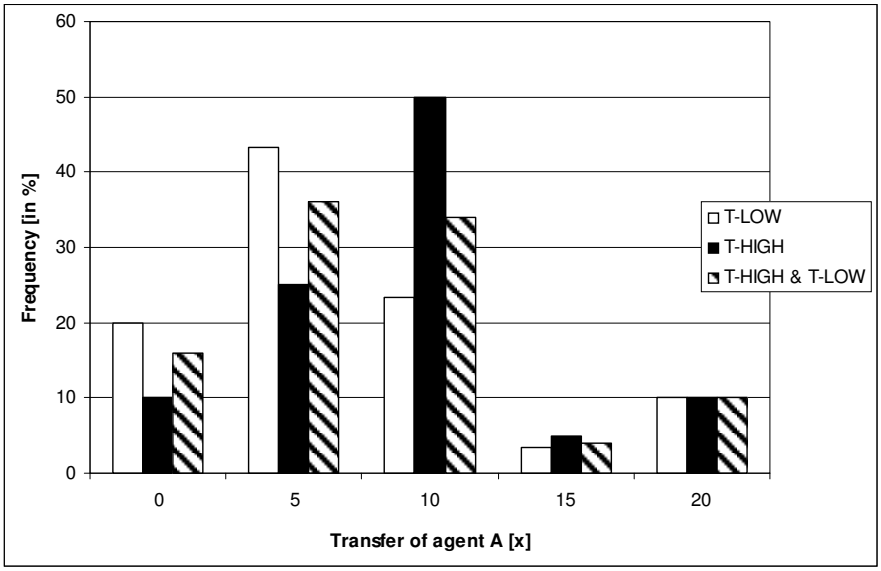
Table 2: Mean and standard deviation of A's transfer

Treatment	Mean of x	Standard deviation of x	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) to agent B. This is considerably larger than 0, but smaller than half of the endowment, which is observed on average in standard trust games in which $q = 1$ (Camerer, 2003, p. 86). In both

[10] The FF-specification completely supports Hypothesis 3. The DK-specification predicts $y(x)$ to be higher in T-LOW than in T-HIGH for $x = 20$, but not necessarily for all $x > 0$.

Figure 2: Distribution of A's transfer



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 more than 60 % of their initial endowment (some even their whole initial endowment). This is considerably different to results from standard dictator games ($q = 0$). In the benchmark treatment by Forsythe et al. (1994) (the paid dictator game conducted in April with a pie of 5 \$) about 55 % of dictators transfer a strictly positive amount, less than 20 % half of their endowment or more, and no dictator transfers more than 60 % of his endowment. This suggests that the distribution of x shifts towards higher values when $q > 0$ compared to $q = 0$.[11] When we consider the distributions of x of the treatments separately, we observe that the distribution in T-HIGH is more to the right than in T-LOW. Taking these observations together it seems that agents A react to differences in q and tend to send more the higher the probability that agent B can reciprocate.

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.

[11] 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 centered on higher values than the distribution of transfers in the corresponding trust game with $q = 0$.

Table 3: Mean and standard deviation of B's return transfer

Treatment	x	Mean of y(x)	Standard deviation of y(x)	Mean of 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).[12] Table 3 also reports the average return transfer divided by the transfer. Independent of $x > 0$ it is below 1. Hence, a strictly positive transfer does not pay off for agent A on average, even if agent A knew beforehand that the game is not stopped. Nevertheless, the mean of $y(x)/x$ increases 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

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 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. Nevertheless, Table 4 shows that there are some agents B that return nothing given $x > 0$. The percentage of these observations decreases in x. Still, 25 % of agents B

[12] 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 tests on Hypothesis 1

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

in T-HIGH and 20 % of agents B in T-LOW return nothing given $x = 20$.

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 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.

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.

We summarize our findings in the following two results:

Result 1: Rewards for actions that seem altruistic

Similar to previous studies on trust games (see Camerer, 2003, p. 86), we observe that agent B returns significantly positive amounts. On average, these amounts increase in agent A's transfer.

Result 2: No depreciation for the expectation of future rewards

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

These results are consistent with the predictions of models of social preferences based on outcomes but inconsistent with the predictions of the self-interest model. Our models of intention-based reciprocity may predict no treatment differences for agent B, but only for individuals that are sufficiently insensitive to reciprocity concerns and return $y(x) = 0$ in both treatments, or for individuals that are extremely sensitive to reciprocity concerns and return $y(x) = 3 * x$ in both treatments. In our data there is no evidence for individuals that are extremely sensitive to reciprocity concerns. There are individuals that seem to be sufficiently insensitive to reciprocity concerns and return nothing (cf. Table 4). On average, though, agents B return strictly positive amounts. Consequently, the predictions of our models of intention-based reciprocity are inconsistent with our aggregate results.

We conclude that the kindness of an action and the action's 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 formulate possible explanations for our findings and evaluate them with questionnaire data.

4.3.1 Incorrect higher order beliefs of agent B

Assume agent B does not expect agent A to expect future rewards. Then, agent B does not depreciate agent A's transfer. From other experimental studies we know that individuals have difficulties to draw inferences from other individual's actions

and correctly form beliefs (e.g. Anderson and Holt, 1997; Hung and Plott, 2001; Kariv, 2005; Nöth and Weber, 2003; Goeree et al., 2007). This is why we elicited agent B's actual second order beliefs and analyze their effect on his decision.[13] 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 actual expectation of agent A's expected future returns for a given x). First, we estimate an OLS regression. Second, we run a two-stage least squares instrumental variable regression in which we instrument for the product of agent B's second order belief with q for a given x. The instrument we use is q itself since it is exogenous and, by definition, correlated with the instrumented variable. We run the second regression because agent B's second order belief for x may be endogenous and, therefore, our estimated OLS coefficient could be biased and inconsistent. Table 6 presents the results of our regressions for $x > 0$.[14]

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

Dependent variable: y(x)	OLS-c	2SLS-IV-c
Intercept	- 03.05***	- 01.67
x	+00.79***	+00.33
Agent B's second order belief * q	+00.24	+00.77
Number of observations	124	124
R-squared	0.3384	0.2700

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

In both regressions agent B's actual belief about agent A's expected return does not significantly affect or spoil agent B's return. In OLS-c the only significant regressor is agent A's transfer: The higher agent A's transfer, the more agent B returns.[15]

[13] Agent B's second order beliefs were elicited in a non-incentivized way after agent B 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 beliefs significantly differ from those elicited by agents B who did not decide upon $y(x)$ because the lottery stopped the game after agent A's decision. We run 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.

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

[15] These results do not qualitatively change if we control for sex, age, and subject of studies.

We conclude that incorrect higher order beliefs of agent B seem not to be a plausible explanation for why an actor's expectation of future rewards does not spoil his gift.

4.3.2 Effect on the perception and on emotions only

An actor's expectation to receive future rewards may only spoil the reactor's perception of the gift or the reactor's emotions without affecting the actual reaction to the gift. Table 7 reports one-sided p-values of Mann-Whitney-U tests on whether (i) a gift is perceived more kind in T-LOW, (ii) negative emotions are less strongly experienced in T-LOW, and (iii) positive emotions are more strongly experienced in T-LOW.[16]

For all $x \in \{0, 5, 10, 15, 20\}$ the average perceived kindness is higher in T-LOW and negative (positive) emotions are less (more) strongly pronounced on average in T-LOW. However, only for some x treatment differences are significant. While there are hardly any significant differences for the perceived kindness and gladness, we observe a few for appreciation and negative emotions like anger and contempt.

We conclude that agent A's intentions may affect agent B's emotions and perception of agent A's action. This effect, however, does not seem to carry over 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 the action's rewards are not spoiled by the actor's expectation of future rewards in our setting. One reason may be that agent B voluntarily decides 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 involuntarily given. Models of intention-based reciprocity do not account for this consideration.

[16] In our questionnaire agents B indicated on a scale ranging from 1 to 7 how kind they perceive a given transfer by agent A. 1 represented "very unkind", while 7 represented "very kind". Furthermore, they indicated on a scale ranging from 1 to 7 with which intensity they hypothetically sensed an emotion for each x. If they did not sense an emotion at all, they were asked to indicate 1 for this particular emotion and given x.

Table 7: One-sided p-values of Mann-Whitney-U tests on kindness and emotions

Attribute	x	Mean in T-HIGH	Mean in T-LOW	p-value
Perceived	0	1.75	2.33	0.1058
kindness	5	2.88	3.27	0.0964
	10	4.00	4.67	0.0284
	15	5.56	5.60	0.2567
	20	6.38	6.47	0.2294
Anger	0	4.44	3.67	0.1449
	5	3.50	2.53	0.0669
	10	2.75	1.47	0.0154
	15	2.00	1.13	0.0023
	20	1.25	1.00	0.0820
Contempt	0	4.69	3.40	0.0358
	5	3.56	2.53	0.0384
	10	2.88	1.60	0.0079
	15	2.25	1.47	0.0097
	20	1.88	1.40	0.4185
Gladness	0	1.19	1.73	0.2011
	5	3.06	3.53	0.1242
	10	4.31	4.80	0.0989
	15	5.56	5.73	0.3273
	20	6.56	6.87	0.1791
Appreciation	0	1.81	2.40	0.1439
	5	2.56	3.67	0.0054
	10	3.63	4.80	0.0060
	15	4.50	5.33	0.0579
	20	5.44	5.80	0.3954

Number of observations in T-HIGH: 16, in T-LOW: 15

Another reason may be that kindness is not an absolute measure but a relative one that considers the ranking of actions for a *given* action set. $x = 20$, for example, may be perceived as the kindest action of agent A's action set in a treatment 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 the action's rewards are reduced by the actor's expectation to receive future rewards.

Our results suggest that behavior that seems altruistic is rewarded. The more altruistic it seems, the higher is the reward in return. In our experiment second movers in a modified trust game return significantly positive rewards to first movers. On average, these rewards increase in the first mover's transfer. The reward for a given action, however, does not vary in the actor's expectation to receive future rewards. We observe in the modified trust game that return transfers do not significantly vary in the probability that the second mover can reciprocate. 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. These observations are consistent with the predictions of models of social preferences based on outcomes but inconsistent with the predictions of the self-interest model and our models of intention-based reciprocity. Hence, individuals in this setting seem to condition their behavior on outcomes rather than on intentions.

These results seem to be relevant for political as well as commercial campaigns, which 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.[17] Therefore, workers 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 also contributes to the discussion of intentions. Our results suggest that a specific kind of intentions, namely to expect a return, does not significantly affect the reaction to an action that seems altruistic. Of course, intentions may well be crucial for other sorts of behavior, e.g. for the reaction to socially undesired behavior. Intentions play an important role in criminal law. Hence, the effect of intentions may depend on the specific context.

[17] A worker's wage may, for instance, depend 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 answered all control questions correctly, participants were informed about their role in the experiment and we proceeded to the decision stages. First, agents A decided upon x. Second, a computer program determined randomly which games of a session were stopped. Each game of a session had the same probability that it is stopped, which corresponded to q of the implemented treatment of the session. Third, agents B were informed about whether the 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 what they would transfer in return for each x if the game continued. After the participants made their decisions, they were asked questions whose answers were not related to any payment, e.g. agents A were asked how many points they believe agent B transfers in return for each x given the game is not stopped, and agents B were asked 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 sex, 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 were originally written in German. The translated instructions for T-HIGH can be found in the following. The instructions for T-LOW are the same except that the probability that the game continues right after agent A's decision is

50 % (instead of 80 %).

6.1.2 Translated instructions for 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 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**.

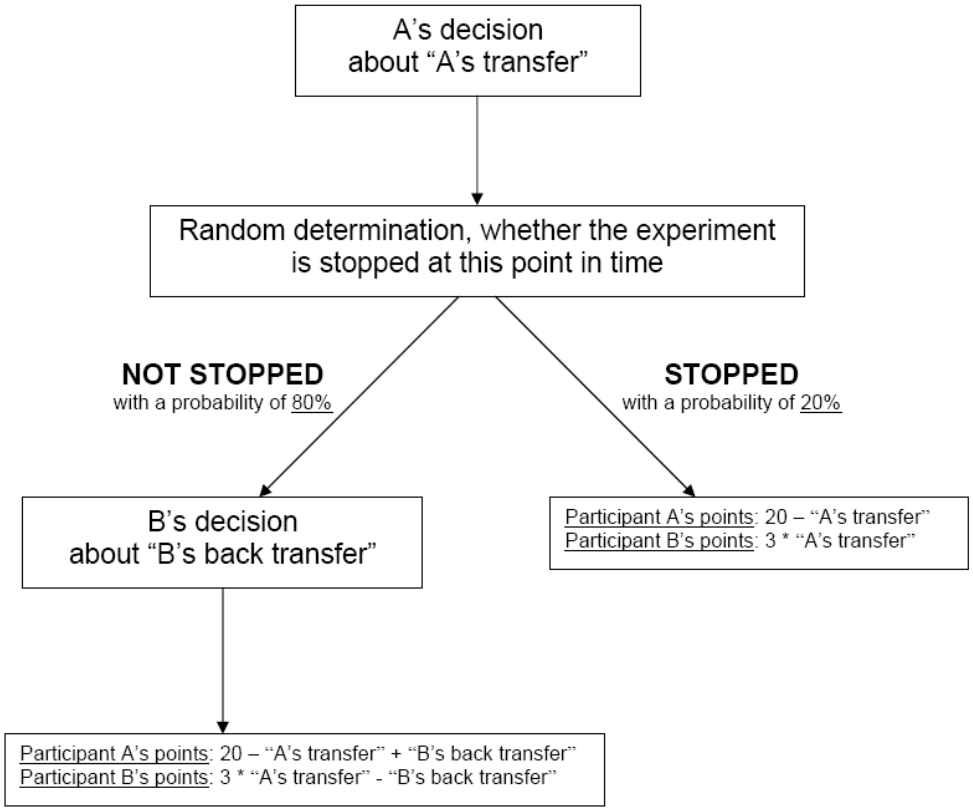
In case the experiment is not stopped right after participant A's decision, participant B makes the decision about the back transfer. In order to do that, participant B indicates for each possible transfer of participant A his selected amount on the following computer screen and presses the OK-button. 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 DK-specification

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 expectation of his own material payoff that depends on his strategy and his belief about individual j's strategy, $Y_i \geqslant 0$ 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, whereas π_i, κ_i, and λ_i depend on individual i's strategy, individual i's belief about individual j's strategy, and 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 material payoff 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 by varying his strategy.[18] λ_i is defined as individual i's belief about individual j's expectation of individual i's material payoff 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 by varying his strategy.

Note that an individual's beliefs are updated in the course of the game and, therefore, may differ after different histories of play. Updated beliefs after a given history equal initial beliefs, except for the choices that were already made and lead to the given history. Updated beliefs assign a probability of 1 to already made choices. Consider, for example, individual i that initially believes individual j to play action a with probability p and action b with probability $1 - p$ (which may, indeed, be correct). After individual j's action a has realized, individual i believes that individual j has chosen a with probability 1 (and not p). As beliefs are updated, also 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.

[18] Dufwenberg and Kirchsteiger (2004) define the reference payoff in a more general way that is equivalent to our notion in our setup.

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.

6.2.2 Our specification

Dufwenberg and Kirchsteiger (2004) restrict attention to finite multi-stage games without nature. For our context, we could simply use their framework and 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. This, however, leads 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 initial belief about agent A's strategy, nature's strategy, and agent A's belief about 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 has chosen the given transfer (with probability 1), that nature has chosen to continue the game (with probability 1), and that agent A believes that nature has chosen 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 believes that nature has chosen 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 to avoid that, we undertake a small and natural modification of the way how agent B perceives the kindness of agent A's strategy in the course of the game. At agent B's decision node we let him evaluate the kindness of agent A's strategy on the basis of his belief that agent A believes that nature has chosen to continue the game with probability 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 $y(x) \in [0, 3 * x]$.[19] At his decision node he believes that agent A has chosen x (with probability 1), that nature has chosen to continue the game (with probability 1), and that agent A believes that agent B returns

[19] Here and in the following we focus on agent B's pure strategies only.

$\widetilde{y}(0) = 0$, $\widetilde{y}(5) \in [0, 15]$, $\widetilde{y}(10) \in [0, 30]$, $\widetilde{y}(15) \in [0, 45]$, $\widetilde{y}(20) \in [0, 60]$, where $\widetilde{y}(x)$ represents agent B's second order belief for x.

Then, agent B's expectation of his own material payoff is equal to

$$\pi_B\left(y(x), x\right) = 3 * x - y(x),$$

and agent B's perception of the kindness of his own strategy, $y(x)$, is equal to

$$\kappa_B\left(y(x), x\right) = 20 - x + y(x) - ref_{\kappa_B}\left(x\right), \text{ with } ref_{\kappa_B}\left(x\right) = \tfrac{(20-x+0)+(20-x+3*x)}{2}.$$

The first term of $\kappa_B\left(y(x), x\right)$, $20 - x + y(x)$, refers to agent B's expectation of agent A's material payoff, while the second, $ref_{\kappa_B}\left(x\right)$, to the corresponding reference payoff that is the mean of the minimum he (believes he) can assign to agent A with $y(x) \in [0, 3 * x]$, $20 - x + 0$, and its maximum, $20 - x + 3 * x$.

Furthermore, agent B's perception of the kindness of agent A's strategy, x, is equal to

$$\lambda_B\left(\widetilde{y}(\cdot), x\right) = 3 * x - q * \widetilde{y}(x) - ref_{\lambda_B}\left(\widetilde{y}(\cdot)\right).$$

The first term of $\lambda_B\left(\widetilde{y}(\cdot), x\right)$, $3 * x - q * \widetilde{y}(x)$, represents agent B's belief about agent A's expectation of agent B's material payoff, which depends on agent B's belief about agent A's action, x, as well as agent B's belief about agent A's belief about agent B's strategy, $\widetilde{y}(x)$, and nature's move. The second term of λ_B, $ref_{\lambda_B}\left(\widetilde{y}(\cdot)\right)$, represents the corresponding reference payoff that is the mean of the minimum agent B believes agent A believes he (agent A) can assign to agent B with $x \in \{0, 5, 10, 15, 20\}$ and its maximum. As $\widetilde{y}(x) \in [0, 3 * x]$, the minimum of $3 * x - q * \widetilde{y}(x)$ is equal to 0 which is attained at $x = 0$. The maximum of $3 * x - q * \widetilde{y}(x)$ depends on agent B's second order beliefs $\widetilde{y}(x)$ for all x. Note that it is not necessarily equal to $3 * 20 - q * \widetilde{y}(20)$ which is attained at $x = 20$.

Hence, agent B's utility function is the following

$$U_B\left(y(x), x, \widetilde{y}(\cdot)\right) = \pi_B\left(y(x), x\right) + Y_B * \kappa_B\left(y(x), x\right) * \lambda_B\left(\widetilde{y}(\cdot), x\right) =$$
$$3 * x - y(x) + Y_B * \left(y(x) - \tfrac{3*x}{2}\right) * \left(3 * x - q * \widetilde{y}(x) - ref_{\lambda_B}\left(\widetilde{y}(\cdot)\right)\right).$$

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 $y(x) \in [0, 3 * x]$ for all $x \in \{0, 5, 10, 15, 20\}$.

1. $y(x)$ **(weakly) increases in** $x \; \forall \; q \in (0, 1)$**.**

Suppose not. Then there exist $x', x \in \{0, 5, 10, 15, 20\}$ with $x' > x$ but $y(x') < y(x)$. As in any SRE initial beliefs about strategies are correct, e.g. $y(x) = \widetilde{y}(x)$ and $y(x') = \widetilde{y}(x')$, agent B believes that agent A intends to assign him more expected payoff with x' than with x since $3 * x' - y(x') * q > 3 * x - y(x) * q$. As $ref_{\lambda_B}(\widetilde{y}(\cdot))$ is the same under x' and x, we have $\lambda_B(\widetilde{y}(\cdot), x') > \lambda_B(\widetilde{y}(\cdot), x)$, i.e. agent B perceives strategy x' as kinder than strategy x. Nevertheless, agent B returns less when he receives $3 * x'$ than when he receives $3 * x$. From revealed preferences it must be the case that:

$$U_B(y(x'), x', \widetilde{y}(\cdot)) \geq U_B(y(x), x', \widetilde{y}(\cdot))$$

and

$$U_B(y(x), x, \widetilde{y}(\cdot)) \geq U_B(y(x'), x, \widetilde{y}(\cdot))$$

because $y(x)$ is available given x' (since $y(x) \leq 3 * x < 3 * x'$), and $y(x')$ is available given x (since $y(x') < y(x) \leq 3 * x$). The two inequalities can be written as

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

and

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

which can be rewritten as

$$\tfrac{1}{Y_B} \geq \lambda_B(\widetilde{y}(\cdot), x') \text{ and } \tfrac{1}{Y_B} \leq \lambda_B(\widetilde{y}(\cdot), x).$$

This implies $\lambda_B(\widetilde{y}(\cdot), x) \geq \lambda_B(\widetilde{y}(\cdot), x')$ which is a contradiction.

2. $(3 * x - y(x) * q)$ **(weakly) increases in** $x \ \forall \ q \in (0, 1)$**.**

Suppose not. Then, there exist $x', x \in \{0, 5, 10, 15, 20\}$ with $x' > x$ but $(3 * x' - y(x') * q) < (3 * x - y(x) * q)$. As in any SRE initial beliefs about strategies are correct, e.g. $y(x) = \widetilde{y}(x)$ and $y(x') = \widetilde{y}(x')$, agent B believes that agent A intends so assign him less payoff with x' than with x. As $ref_{\lambda_B}(\widetilde{y}(\cdot))$ is the same under x' and x, we have $\lambda_B(\widetilde{y}(\cdot), x') < \lambda_B(\widetilde{y}(\cdot), x)$, i.e. agent B perceives strategy x as kinder than strategy x'. From revealed preferences it must be the case that:

$$U_B\left(y(x'), x', \widetilde{y}(\cdot)\right) \geq U_B\left(y(x), x', \widetilde{y}(\cdot)\right)$$

and

$$U_B\left(y(x), x, \widetilde{y}(\cdot)\right) \geq U_B\left(\left(y(x') - 3 * (x' - x)\right), x, \widetilde{y}(\cdot)\right)$$

because $y(x)$ is available given x' (since $y(x) \leq 3 * x < 3 * x'$), and $y(x') - 3 * (x' - x)$ is available given x. The latter is true since $y(x') - 3 * (x' - x) \leq 3 * x$, which is equivalent to $y(x') \leq 3 * x'$, and $y(x') - 3 * (x' - x) \geq 0$ since the above assumed $(3 * x' - y(x') * q) < (3 * x - y(x) * q)$ implies $y(x') - 3 * (x' - x) > y\left(x\right) \geq 0$. The two inequalities can be written as

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

and

$$\begin{gathered} 3 * x - y(x) + Y_B * \left(y(x) - \tfrac{3}{2} * x\right) * \lambda_B\left(\widetilde{y}(\cdot), x\right) \geq \\ 3 * x - \left(y(x') - 3 * (x' - x)\right) + Y_B * \left(\left(y(x') - 3 * (x' - x)\right) - \tfrac{3}{2} * x\right) * \lambda_B\left(\widetilde{y}(\cdot), x\right). \end{gathered}$$

$3 * x' - y(x') * q < 3 * x - y(x) * q$ is equivalent to $3 * (x' - x) < q * (y(x') - y(x))$ and implies (i) $y(x') > y(x)$ and (ii) $3 * (x' - x) < y(x') - y(x)$ which is equivalent to $y(x') - 3 * (x' - x) > y(x)$. Therefore, we can be rewrite the two inequalities as

$$\lambda_B\left(\widetilde{y}(\cdot), x'\right) \geq \tfrac{1}{Y_B} \text{ and } \lambda_B\left(\widetilde{y}(\cdot), x\right) \leq \tfrac{1}{Y_B}.$$

This implies $\lambda_B\left(\widetilde{y}(\cdot), x'\right) \geq \lambda_B\left(\widetilde{y}(\cdot), x\right)$ which is a contradiction.

3. $\lambda_B\left(\widetilde{y}(\cdot), x\right)$ (weakly) increases in x $\forall$ $q \in (0, 1)$.

As in any SRE initial beliefs about strategies are correct, e.g. $y(x) = \widetilde{y}(x)$ for all $x \in \{0, 5, 10, 15, 20\}$, and our second property holds, agent B believes that agent A intends to assign him (weakly) more expected material payoff the higher x. As $ref_{\lambda_B}\left(\widetilde{y}(\cdot)\right)$ is the same for any $x \in \{0, 5, 10, 15, 20\}$, $\lambda_B\left(\widetilde{y}(\cdot), x\right)$ (weakly) increases in x.

4. The higher q the (weakly) smaller $y(x)$ $\forall$ $q \in (0, 1)$ and $x = 20$.

Suppose not. Then, there exist $q', q \in (0, 1)$ with $q' > q$ and an SRE for q' with $y(20)_{q'}$ and an SRE for q with $y(20)_q$ such that $y(20)_{q'} > y(20)_q$. As in any SRE initial beliefs about strategies are correct, agent B believes that agent A intends to assign him more expected payoff with $x = 20$ when the probability that the game

is not stopped is q rather than q' because $3*20-y(20)_q*q>3*20-y(20)_{q'}*q'$. $ref_{\lambda_B}(\widetilde{y}(\cdot))$ may be different for q' and q. Due to our second property and correct initial beliefs about strategies, we can simply calculate $ref_{\lambda_B}(\widetilde{y}(\cdot))$ as the mean of the minimum agent B believes agent A believes he (agent A) can assign to agent B with $x\in\{0,5,10,15,20\}$, which is attained at $x=0$, and its maximum, which is attained at $x=20$. Hence, $ref_{\lambda_B}(\widetilde{y}(\cdot)_q)_q=(3*20-y(20)_q*q)*\frac{1}{2}$ and $ref_{\lambda_B}(\widetilde{y}(\cdot)_{q'})_{q'}=(3*20-y(20)_{q'}*q')*\frac{1}{2}$. As a consequence, agent B perceives $x=20$ as kinder in the SRE with q than in the one with q', i.e. $\lambda_B(\widetilde{y}(\cdot)_q,20)_q>\lambda_B(\widetilde{y}(\cdot)_{q'},20)_{q'}$.[20] This is because $(3*20-y(20)_q*q)*\frac{1}{2}>(3*20-y(20)_{q'}*q')*\frac{1}{2}$. Nevertheless, agent B returns more in the SRE with q' than in the one with q. From revealed preferences it must be the case that:

$$U_B(y(20)_{q'},20,\widetilde{y}(\cdot)_{q'})_{q'}\geq U_B(y(20)_q,20,\widetilde{y}(\cdot)_{q'})_{q'}$$

and

$$U_B(y(20)_q,20,\widetilde{y}(\cdot)_q)_q\geq U_B(y(20)_{q'},20,\widetilde{y}(\cdot)_q)_q$$

because $y(20)_q$ and $y(20)_{q'}$ are both available given $x=20$. The two inequalities can be written as

$$\begin{gathered}3*20-y(20)_{q'}+Y_B*\left(y(20)_{q'}-\tfrac{3}{2}*20\right)*\lambda_B(\widetilde{y}(\cdot)_{q'},20)_{q'}\geq\\ 3*20-y(20)_q+Y_B*\left(y(20)_q-\tfrac{3}{2}*20\right)*\lambda_B(\widetilde{y}(\cdot)_{q'},20)_{q'}\end{gathered}$$

and

$$\begin{gathered}3*20-y(20)_q+Y_B*\left(y(20)_q-\tfrac{3}{2}*20\right)*\lambda_B(\widetilde{y}(\cdot)_q,20)_q\geq\\ 3*20-y(20)_{q'}+Y_B*\left(y(20)_{q'}-\tfrac{3}{2}*20\right)*\lambda_B(\widetilde{y}(\cdot)_q,20)_q\end{gathered}$$

which can be rewritten as

$$\lambda_B(\widetilde{y}(\cdot)_{q'},20)_{q'}\geq\tfrac{1}{Y_B}\text{ and }\lambda_B(\widetilde{y}(\cdot)_q,20)_q\leq\tfrac{1}{Y_B}.$$

This implies $\lambda_B(\widetilde{y}(\cdot)_{q'},20)_{q'}\geq\lambda_B(\widetilde{y}(\cdot)_q,20)_q$ which is a contradiction.

5. If $y(20)_q=y(20)_{q'}$ for $q',q\in(0,1)$ with $q'>q$, then either $y(20)_q=y(20)_{q'}=60$ or $y(20)_q=y(20)_{q'}=0$.

Suppose not. Then, there exist $q',q\in(0,1)$ with $q'>q$ and an SRE for q' with $y(20)_{q'}$ and an SRE for q with $y(20)_q$ such that $0<y(20)_{q'}=y(20)_q<60$. As in any

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

SRE initial beliefs about strategies are correct, agent B believes that agent A intends to assign him more expected payoff with $x = 20$ when the probability that the game is not stopped is q rather than q' because $3 * 20 - y(20)_q * q > 3 * 20 - y(20)_{q'} * q'$ (with $y(20)_{q'} = y(20)_q > 0$). Due to our second property and correct initial beliefs about strategies, we can simply calculate $ref_{\lambda_B}\left(\widetilde{y}(\cdot)_q\right)_q = (3 * 20 - y(20)_q * q) * \frac{1}{2}$ and $ref_{\lambda_B}\left(\widetilde{y}(\cdot)_{q'}\right)_{q'} = (3 * 20 - y(20)_{q'} * q') * \frac{1}{2}$. As a consequence, agent B perceives $x = 20$ as kinder in the SRE with q than in the one with q', i.e. $\lambda_B\left(\widetilde{y}(\cdot)_q, 20\right)_q > \lambda_B\left(\widetilde{y}(\cdot)_{q'}, 20\right)_{q'}$.[21] Nevertheless, agent B returns the same in the SRE with q' as in the one with q. From revealed preferences it must be the case that:

$$U_B\left(y(20)_{q'}, 20, \widetilde{y}(\cdot)_{q'}\right)_{q'} \geq U_B\left(0, 20, \widetilde{y}(\cdot)_{q'}\right)_{q'}$$

and

$$U_B\left(y(20)_q, 20, \widetilde{y}(\cdot)_q\right)_q \geq U_B\left(60, 20, \widetilde{y}(\cdot)_q\right)_q$$

because 0 and 60 are available given $x = 20$. The two inequalities can be written as

$$\begin{gathered} 3 * 20 - y(20)_{q'} + Y_B * \left(y(20)_{q'} - \tfrac{3}{2} * 20\right) * \lambda_B\left(\widetilde{y}(\cdot)_{q'}, 20\right)_{q'} \geq \\ 3 * 20 - 0 + Y_B * \left(0 - \tfrac{3}{2} * 20\right) * \lambda_B\left(\widetilde{y}(\cdot)_{q'}, 20\right)_{q'} \end{gathered}$$

and

$$\begin{gathered} 3 * 20 - y(20)_q + Y_B * \left(y(20)_q - \tfrac{3}{2} * 20\right) * \lambda_B\left(\widetilde{y}(\cdot)_q, 20\right)_q \geq \\ 3 * 20 - 60 + Y_B * \left(60 - \tfrac{3}{2} * 20\right) * \lambda_B\left(\widetilde{y}(\cdot)_q, 20\right)_q \end{gathered}$$

which can be rewritten as

$$\lambda_B\left(\widetilde{y}(\cdot)_{q'}, 20\right)_{q'} \geq \tfrac{1}{Y_B} \text{ and } \lambda_B\left(\widetilde{y}(\cdot)_q, 20\right)_q \leq \tfrac{1}{Y_B}.$$

This implies $\lambda_B\left(\widetilde{y}(\cdot)_{q'}, 20\right)_{q'} \geq \lambda_B\left(\widetilde{y}(\cdot)_q, 20\right)_q$ which is a contradiction.

6. $\lambda_B\left(\widetilde{y}(\cdot), 20\right)$ is (weakly) larger when $q = 0.5$ than when $q = 0.8$.

Suppose not. Then, there exist an SRE for $q = 0.5$ with $y(20)_{0.5}$ and an SRE for $q = 0.8$ with $y(20)_{0.8}$ such that $\lambda_B\left(\widetilde{y}(\cdot)_{0.5}, 20\right)_{0.5} < \lambda_B\left(\widetilde{y}(\cdot)_{0.8}, 20\right)_{0.8}$. Due to our second property and correct initial beliefs about strategies, this implies that $(60 - 0.5 * y(20)_{0.5}) * \frac{1}{2} < (60 - 0.8 * y(20)_{0.8}) * \frac{1}{2}$ and, therefore, $y(20)_{0.5} > y(20)_{0.8}$. From revealed preferences it must be the case that:

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

$$U_B\left(y(20)_{0.5}, 20, \widetilde{y}(\cdot)_{0.5}\right)_{0.5} \geq U_B\left(y(20)_{0.8}, 20, \widetilde{y}(\cdot)_{0.5}\right)_{0.5}$$

and

$$U_B\left(y(20)_{0.8}, 20, \widetilde{y}(\cdot)_{0.8}\right)_{0.8} \geq U_B\left(y(20)_{0.5}, 20, \widetilde{y}(\cdot)_{0.8}\right)_{0.8}$$

because $y(20)_{0.5}$ and $y(20)_{0.8}$ are available given $x = 20$. The two inequalities can be written as

$$\begin{aligned} &3 * 20 - y(20)_{0.5} + Y_B * \left(y(20)_{0.5} - \tfrac{3}{2} * 20\right) * \lambda_B\left(\widetilde{y}(\cdot)_{0.5}, 20\right)_{0.5} \geq \\ &3 * 20 - y(20)_{0.8} + Y_B * \left(y(20)_{0.8} - \tfrac{3}{2} * 20\right) * \lambda_B\left(\widetilde{y}(\cdot)_{0.5}, 20\right)_{0.5} \end{aligned}$$

and

$$\begin{aligned} &3 * 20 - y(20)_{0.8} + Y_B * \left(y(20)_{0.8} - \tfrac{3}{2} * 20\right) * \lambda_B\left(\widetilde{y}(\cdot)_{0.8}, 20\right)_{0.8} \geq \\ &3 * 20 - y(20)_{0.5} + Y_B * \left(y(20)_{0.5} - \tfrac{3}{2} * 20\right) * \lambda_B\left(\widetilde{y}(\cdot)_{0.8}, 20\right)_{0.8} \end{aligned}$$

which can be rewritten as

$$\lambda_B\left(\widetilde{y}(\cdot)_{0.5}, 20\right)_{0.5} \geq \tfrac{1}{Y_B} \text{ and } \lambda_B\left(\widetilde{y}(\cdot)_{0.8}, 20\right)_{0.8} \leq \tfrac{1}{Y_B}.$$

This implies $\lambda_B\left(\widetilde{y}(\cdot)_{0.5}, 20\right)_{0.5} \geq \lambda_B\left(\widetilde{y}(\cdot)_{0.8}, 20\right)_{0.8}$ which is a contradiction.

7. Agent A's expected return from $x = 20$, $q * y(20)$, is (weakly) smaller when $q = 0.5$ than when $q = 0.8$.

Our sixth property states $\lambda_B\left(\widetilde{y}(\cdot)_{0.5}, 20\right)_{0.5} \geq \lambda_B\left(\widetilde{y}(\cdot)_{0.8}, 20\right)_{0.8}$. Due to our second property and correct initial beliefs about strategies, this implies $(60 - 0.5 * y(20)_{0.5}) * \frac{1}{2} \geq (60 - 0.8 * y(20)_{0.8}) * \frac{1}{2}$ which is equivalent to $0.8 * y(20)_{0.8} \geq 0.5 * y(20)_{0.5}$.

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 $y(x) \in [0, 3 * x]$ for all $x \in \{0, 5, 10, 15, 20\}$. In the following we show that at least one such SRE exists for each of our treatments.

Lemma 1: $\forall\, x \in \{0, 5, 10, 15, 20\}$ and $q \in (0, 1)$ there exists an optimal pure action for agent B, $y(x) \in [0, 3 * x]$, such that agent B's initial beliefs about agent A's beliefs about agent B's actions are all correct, i.e. $y(x) = \widetilde{y}(x)$ for all $x \in \{0, 5, 10, 15, 20\}$.

Take an $x \in \{0,5,10,15\}$, a $\widetilde{y}(20) \in [0,60]$, and the fact that $ref_{\lambda_B}(\widetilde{y}(\cdot)) = \frac{60-q*\widetilde{y}(20)+0}{2}$ (as it is the case in all SRE in which agent B chooses a pure strategy due to our second property). Then, agent B's utility function is $U_B(y(x), x, \widetilde{y}(\cdot)) = 3*x-y(x)+Y_B*\big(y(x)-\frac{3*x}{2}\big)*\Big(3*x-q*\widetilde{y}(x)-\frac{60-q*\widetilde{y}(20)+0}{2}\Big)$. As $U_B(y(x), x, \widetilde{y}(\cdot))$ does not depend on $\widetilde{y}(x')$ with $x' \in \{0,5,10,15\}\setminus x$ and x and $\widetilde{y}(20)$ are fixed, we rewrite agent B's utility function as $U_B(y(x), \widetilde{y}(x))$. $U_B(y(x), \widetilde{y}(x))$ is continuous in $y(x)$ and $\widetilde{y}(x)$, and $U_B(\cdot, \widetilde{y}(x))$ is quasi-concave in $\widetilde{y}(x)$. By choosing a $y(x) \in G(\widetilde{y}(x)) = [0, 3*x]$ 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 at least one $y(x) \in [0, 3*x]$ that maximizes $U_B(y(x), \widetilde{y}(x))$ 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 set of $y(x) \in [0, 3*x]$ which maximize $U_B(y(x), \widetilde{y}(x))$ 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. agent B's initial beliefs about agent A's beliefs about agent B's actions for x are correct. We apply Kakutani's Fixed Point Theorem and conclude that at least one fixed point exists.

Now, take $x = 20$, and the fact that $ref_{\lambda_B}(\widetilde{y}(\cdot)) = \frac{60-q*\widetilde{y}(20)+0}{2}$. Then, agent B's utility function is $U_B(y(20), 20, \widetilde{y}(\cdot)) = 3*20 - y(20) + Y_B*\big(y(20)-\frac{3*20}{2}\big)*\Big(3*20-q*\widetilde{y}(20)-\frac{60-q*\widetilde{y}(20)+0}{2}\Big)$. As $U_B(y(20), 20, \widetilde{y}(\cdot))$ does not depend on $\widetilde{y}(x')$ with $x' \in \{0,5,10,15\}$ and x is fixed at 20, we rewrite agent B's utility function as $U_B(y(20), \widetilde{y}(20))$. Again, $U_B(y(20), \widetilde{y}(20))$ is continuous in $y(20)$ and $\widetilde{y}(20)$, $U_B(\cdot, \widetilde{y}(20))$ is quasi-concave, and $[0,60]$, the attainable set of pure actions, 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 at least one $y(20) \in [0,60]$ that maximizes $U_B(y(20), \widetilde{y}(20))$ and that there exist at least one $\widetilde{y}(20)$ that is correct.

From our second property we know that if (i) agent B has some $ref_{\lambda_B}(\widetilde{y}(\cdot))$, which is the same under all $x \in \{0,5,10,15,20\}$, and (ii) his initial beliefs are correct, e.g. $y(x) = \widetilde{y}(x)$ for all $x \in \{0,5,10,15,20\}$, and (iii) he behaves rational in the sense that he chooses an action when its derived utility is (weakly) highest, then $3*x - \widetilde{y}(x)*q$ increases in x and $ref_{\lambda_B}(\widetilde{y}(\cdot)) = \frac{60-q*\widetilde{y}(20)+0}{2}$.

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

Due to Lemma 1, it remains to show that given agent B's pure optimal strategy

agent A has an optimal (possibly randomized) strategy a that is correctly expected by initial beliefs, i.e. $a = \widetilde{a}$ with $\widetilde{a}$ as agent A's initial second order belief on a.

Take any optimal pure strategy of agent B $y(x)$ for all $x \in \{0, 5, 10, 15, 20\}$ which is correctly expected by agent A. Then, agent A's utility function is $U_A(a, y(\cdot), \widetilde{a}) = \pi_A(a, y(\cdot)) + Y_A * \kappa_A(a, y(\cdot)) * \lambda_A(y(\cdot), \widetilde{a})$. Let us define $E(x)$ and $\widetilde{E}(x)$ as the mean of x resulting with strategy a and $\widetilde{a}$, respectively, and $E(y(x))$ and $\widetilde{E}(y(x))$ as the mean of $y(x)$ resulting with strategy a and $\widetilde{a}$, respectively. Then, $\pi_A(a, y(\cdot)) = 20 - E(x) - q * E(y(x))$, $\kappa_A(a, y(\cdot)) = 3 * E(x) - q * E(y(x)) - \frac{0+3*20-q*y(20)}{2}$, and $\lambda_A(y(\cdot), \widetilde{a}) = 20 - \widetilde{E}(x) + q * \widetilde{E}(y(x)) - \frac{20-\widetilde{E}(x)+q*0+20-\widetilde{E}(x)+q*3*\widetilde{E}(x)}{2}$. Hence, $U_A(a, y(\cdot), \widetilde{a}) = 20 - E(x) - q * E(y(x)) + Y_A * \left(3 * E(x) - q * E(y(x)) - \frac{3*20-q*y(20)}{2}\right) * \left(q * \widetilde{E}(y(x)) - \frac{q*3*\widetilde{E}(x)}{2}\right)$. As $y(\cdot)$ is fixed, we can rewrite agent A's utility function as $U_A(a, \widetilde{a})$. $U_A(a, \widetilde{a})$ is continuous in a and $\widetilde{a}$, $U_A(\cdot, \widetilde{a})$ is quasi-concave, and agent A's set of possibly randomized strategies 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 strategies $X^*(\widetilde{a})$ out of which each strategy 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, and 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 FF-specification

6.3.1 Our specification

We consider the same model as for the DK-specification but define κ_i and λ_i differently. The interpretation of these terms, though, remains the same. The reference payoff used for κ_i is equal to individual i's expectation of his own material payoff, π_i, while the reference payoff used for λ_i is equal to individual i's belief about individual j's expectation of individual j's material payoff. Everything else remains the same.

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

In comparison to agent B's corresponding utility function in the DK-specification, $\kappa_B(y(x), x)$ and $\lambda_B(\widetilde{y}(x), x)$ change. Now,

$$\kappa_B(y(x), x) = 20 - x + y(x) - (3 * x - y(x))$$

and

$$\lambda_B\left(\widetilde{y}(x), x\right) = 3 * x - q * \widetilde{y}(x) - (20 - x + q * \widetilde{y}(x)).$$

Hence, agent B's utility function is the following

$$U_B\left(y(x), x, \widetilde{y}(x)\right) =$$
$$3 * x - y(x) + Y_B * (20 - 4 * x + 2 * y\left(x\right)) * (4 * x - 2 * q * \widetilde{y}(x) - 20).$$

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 $y(x) \in [0, 3 * x]$ for all $x \in \{0, 5, 10, 15, 20\}$.

1. $y(x)$ **(weakly) increases in** $x \ \forall \ q \in (0, 1)$**.**

Suppose not. Then there exist $x', x \in \{0, 5, 10, 15, 20\}$ with $x' > x$ but $y(x') < y(x)$. As in any SRE initial beliefs about strategies are correct, e.g. $y(x) = \widetilde{y}(x)$ and $y(x') = \widetilde{y}(x')$, $\lambda_B\left(\widetilde{y}(x'), x'\right) > \lambda_B\left(\widetilde{y}(x), x\right)$ since $4 * x' - 2 * q * y(x') - 20 > 4 * x - 2 * q * y(x) - 20$. Nevertheless, agent B returns less when he receives $3 * x'$ than when he receives $3 * x$. From revealed preferences it must be the case that:

$$U_B\left(y(x'), x', \widetilde{y}(x')\right) \geq U_B\left(y(x), x', \widetilde{y}(x')\right)$$

and

$$U_B\left(y(x), x, \widetilde{y}(x)\right) \geq U_B\left(y(x'), x, \widetilde{y}(x)\right)$$

because $y(x)$ is available given x' (since $y(x) \leq 3 * x < 3 * x'$), and $y(x')$ is available given x (since $y(x') < y(x) \leq 3 * x$). The two inequalities can be written as

$$3 * x' - y(x') + Y_B * (20 - 4 * x' + 2 * y\left(x'\right)) * \lambda_B\left(\widetilde{y}(x'), x'\right) \geq$$
$$3 * x' - y(x) + Y_B * (20 - 4 * x' + 2 * y\left(x\right)) * \lambda_B\left(\widetilde{y}(x'), x'\right)$$

and

$$3 * x - y(x) + Y_B * (20 - 4 * x + 2 * y\left(x\right)) * \lambda_B\left(\widetilde{y}(x), x\right) \geq$$
$$3 * x - y(x') + Y_B * (20 - 4 * x + 2 * y\left(x'\right)) * \lambda_B\left(\widetilde{y}(x), x\right)$$

which can be rewritten as

$$\tfrac{1}{2 * Y_B} \geq \lambda_B\left(\widetilde{y}(x'), x'\right) \text{ and } \tfrac{1}{2 * Y_B} \leq \lambda_B\left(\widetilde{y}(x), x\right).$$

This implies $\lambda_B\left(\widetilde{y}(x),x\right) \geq \lambda_B\left(\widetilde{y}(x'),x'\right)$ which is a contradiction.

2. $\lambda_B\left(\widetilde{y}(x),x\right)$ **(weakly) increases in** $x \ \forall \ q \in \{0.5,0.8\}$**.**

Suppose not. Then, there exist $x', x \in \{0,5,10,15,20\}$ with $x' > x$ but $\lambda_B\left(\widetilde{y}(x'),x'\right) < \lambda_B\left(\widetilde{y}(x),x\right)$. As in any SRE initial beliefs about strategies are correct, e.g. $y(x) = \widetilde{y}(x)$ and $y(x') = \widetilde{y}(x')$, this implies $4 * x' - 2 * q * y(x') - 20 < 4 * x - 2 * q * y(x) - 20$ which is equivalent to $4 * (x' - x) < 2 * q * (y(x') - y(x))$ and implies $y(x') > y(x)$. Furthermore, for $q \in \{0.5,0.8\}$ $y(x) < 3 * x$ in SRE. If not and $y(x) = \widetilde{y}(x) = 3 * x$, $\lambda_B\left(\widetilde{y}(x),x\right) = 4 * x - 2 * 3 * x * q - 20$, which is equal or smaller than 0 for $q \in \{0.5,0.8\}$. Given $\lambda_B\left(\widetilde{y}(x),x\right) \leq 0$, agent B preferred to return nothing instead of $y(x) = 3 * x$.

From revealed preferences it must be the case that:

$$U_B\left(y(x'),x',\widetilde{y}(x')\right) \geq U_B\left(y(x),x',\widetilde{y}(x')\right)$$

and

$$U_B\left(y(x),x,\widetilde{y}(x)\right) \geq U_B\left(3 * x,x,\widetilde{y}(x)\right)$$

because $y(x)$ is available given x' (since $y(x) < 3 * x < 3 * x'$), and $3 * x$ is available given x. The two inequalities can be written as

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

and

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

As $y(x') > y(x)$ and $y(x) < 3 * x$, the two inequalities can be rewritten as

$$\lambda_B\left(\widetilde{y}(x'),x'\right) \geq \tfrac{1}{2 * Y_B} \text{ and } \lambda_B\left(\widetilde{y}(x),x\right) \leq \tfrac{1}{2 * Y_B}.$$

This implies $\lambda_B\left(\widetilde{y}(x'),x'\right) \geq \lambda_B\left(\widetilde{y}(x),x\right)$ which is a contradiction.

3. The higher q **the (weakly) smaller** $y(x) \ \forall \ q \in (0,1)$ **and** $x \in \{0,5,10,15,20\}$**.**

Suppose not. Then, there exist $q', q \in (0,1)$ with $q' > q$ and an SRE for q' with $y(x)_{q'}$ and an SRE for q with $y(x)_q$ such that $y(x)_{q'} > y(x)_q$. As in any SRE initial beliefs about strategies are correct, e.g. $y(x)_{q'} = \widetilde{y}(x)_{q'}$ and $y(x)_q = \widetilde{y}(x)_q$, $\lambda_B\left(\widetilde{y}(x)_q,x\right)_q > \lambda_B\left(\widetilde{y}(x)_{q'},x\right)_{q'}$ since $4 * x - 2 * q * y(x)_q - 20 > 4 * x - 2 * q' * y(x)_{q'} - 20$. Nevertheless, agent B returns more in the SRE with q' than in the one with q for x. From revealed preferences it must be the case that:

$$U_B\left(y(x)_{q'}, x, \widetilde{y}(x)_{q'}\right)_{q'} \geq U_B\left(y(x)_q, x, \widetilde{y}(x)_{q'}\right)_{q'}$$

and

$$U_B\left(y(x)_q, x, \widetilde{y}(x)_q\right)_q \geq U_B\left(y(x)_{q'}, x, \widetilde{y}(x)_q\right)_q$$

because $y(x)_q$ and $y(x)_{q'}$ are both available given x. The two inequalities can be written as

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

and

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

which can be rewritten as

$$\lambda_B\left(\widetilde{y}(x)_{q'}, x\right)_{q'} \geq \tfrac{1}{2*Y_B} \text{ and } \lambda_B\left(\widetilde{y}(x)_q, x\right)_q \leq \tfrac{1}{2*Y_B}.$$

This implies $\lambda_B\left(\widetilde{y}(x)_{q'}, x\right)_{q'} \geq \lambda_B\left(\widetilde{y}(x)_q, x\right)_q$ which is a contradiction.

4. For $x \in \{5, 10, 15, 20\}$ it holds that if $y(x)_q = y(x)_{q'}$ for $q', q \in (0, 1)$ with $q' > q$, then either $y(x)_q = y(x)_{q'} = 3*x$ or $y(x)_q = y(x)_{q'} = 0$.

Suppose not. Then, there exist $q', q \in (0, 1)$ with $q' > q$ and an SRE for q' with $y(x)_{q'}$ and an SRE for q with $y(x)_q$ such that $0 < y(x)_{q'} = y(x)_q < 3*x$. As in any SRE initial beliefs about strategies are correct, $\lambda_B\left(\widetilde{y}(x)_q, x\right)_q > \lambda_B\left(\widetilde{y}(x)_{q'}, x\right)_{q'}$ since $4*x - 2*q*y(x)_q - 20 > 4*x - 2*q'*y(x)_{q'} - 20$. Nevertheless, agent B returns the same in the SRE with q' as in the one with q. From revealed preferences it must be the case that:

$$U_B\left(y(x)_{q'}, x, \widetilde{y}(x)_{q'}\right)_{q'} \geq U_B\left(0, x, \widetilde{y}(x)_{q'}\right)_{q'}$$

and

$$U_B\left(y(x)_q, x, \widetilde{y}(x)_q\right)_q \geq U_B\left(3*x, x, \widetilde{y}(x)_q\right)_q,$$

because 0 and $3*x$ are available given x. The two inequalities can be written as

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

and

$$3 * x - y(x)_q + Y_B * (20 - 4 * x + 2 * y(x)_q) * \lambda_B \left(\widetilde{y}(x)_q, x\right)_q \geq$$
$$3 * x - 3 * x + Y_B * (20 - 4 * x + 2 * 3 * x) * \lambda_B \left(\widetilde{y}(x)_q, 20\right)_q$$

which can be rewritten as

$$\lambda_B \left(\widetilde{y}(x)_{q'}, x\right)_{q'} \geq \tfrac{1}{2 * Y_B} \text{ and } \lambda_B \left(\widetilde{y}(x)_q, x\right)_q \leq \tfrac{1}{2 * Y_B}.$$

This implies $\lambda_B \left(\widetilde{y}(x)_{q'}, x\right)_{q'} \geq \lambda_B \left(\widetilde{y}(x)_q, x\right)_q$ which is a contradiction.

5. $\lambda_B \left(\widetilde{y}(x), x\right)$ is (weakly) larger when $q = 0.5$ than when $q = 0.8$.

Suppose not. Then, there exist an SRE for $q = 0.5$ with $y(x)_{0.5}$ and an SRE for $q = 0.8$ with $y(x)_{0.8}$ such that $\lambda_B \left(\widetilde{y}(x)_{0.5}, x\right)_{0.5} < \lambda_B \left(\widetilde{y}(x)_{0.8}, x\right)_{0.8}$. Due to correct initial beliefs about strategies, this implies that $4 * x - 2 * 0.5 * y(x)_{0.5} - 20 < 4 * x - 2 * 0.8 * y(x)_{0.8} - 20$ and, therefore, $y(x)_{0.5} > y(x)_{0.8}$. From revealed preferences it must be the case that:

$$U_B \left(y(x)_{0.5}, x, \widetilde{y}(x)_{0.5}\right)_{0.5} \geq U_B \left(y(x)_{0.8}, x, \widetilde{y}(x)_{0.5}\right)_{0.5}$$

and

$$U_B \left(y(x)_{0.8}, x, \widetilde{y}(x)_{0.8}\right)_{0.8} \geq U_B \left(y(x)_{0.5}, x, \widetilde{y}(x)_{0.8}\right)_{0.8},$$

because $y(x)_{0.5}$ and $y(x)_{0.8}$ are available given x. The two inequalities can be written as

$$3 * x - y(x)_{0.5} + Y_B * (20 - 4 * x + 2 * y(x)_{0.5}) * \lambda_B \left(\widetilde{y}(x)_{0.5}, x\right)_{0.5} \geq$$
$$3 * x - y(x)_{0.8} + Y_B * (20 - 4 * x + 2 * y(x)_{0.8}) * \lambda_B \left(\widetilde{y}(x)_{0.5}, x\right)_{0.5}$$

and

$$3 * x - y(x)_{0.8} + Y_B * (20 - 4 * x + 2 * y(x)_{0.8}) * \lambda_B \left(\widetilde{y}(x)_{0.8}, x\right)_{0.8} \geq$$
$$3 * x - y(x)_{0.5} + Y_B * (20 - 4 * x + 2 * y(x)_{0.5}) * \lambda_B \left(\widetilde{y}(x)_{0.8}, x\right)_{0.8}$$

which can be rewritten as

$$\lambda_B \left(\widetilde{y}(x)_{0.5}, x\right)_{0.5} \geq \tfrac{1}{2 * Y_B} \text{ and } \lambda_B \left(\widetilde{y}(x)_{0.8}, x\right)_{0.8} \leq \tfrac{1}{2 * Y_B}.$$

This implies $\lambda_B \left(\widetilde{y}(x)_{0.5}, x\right)_{0.5} \geq \lambda_B \left(\widetilde{y}(x)_{0.8}, x\right)_{0.8}$ which is a contradiction.

6. Agent A's expected return from x, $q * y(x)$, is (weakly) smaller when $q = 0.5$ than when $q = 0.8$.

Our fifth property states $\lambda_B \left(\widetilde{y}(x)_{0.5}, x\right)_{0.5} \geq \lambda_B \left(\widetilde{y}(x)_{0.8}, x\right)_{0.8}$. Due to correct initial beliefs about strategies, this implies $4 * x - 2 * 0.5 * y(x)_{0.5} - 20 \geq 4 * x - 2 * 0.8 * y(x)_{0.8} - 20$ which is equivalent to $0.8 * y(x)_{0.8} \geq 0.5 * y(x)_{0.5}$.

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 $y(x) \in [0, 3 * x]$ for all $x \in \{0, 5, 10, 15, 20\}$. In the following we show that at least one such SRE exists for each of our treatments.

Lemma 1': $\forall\ x \in \{0, 5, 10, 15, 20\}$ and $q \in (0, 1)$ there exists an optimal pure action for agent B, $y(x) \in [0, 3 * x]$, such that agent B's initial beliefs about agent A's beliefs about agent B's actions are all correct, i.e. $y(x) = \widetilde{y}(x)$ for all $x \in \{0, 5, 10, 15, 20\}$.

Take an $x \in \{0, 5, 10, 15, 20\}$. Then, agent B's utility function is $U_B\left(y(x), x, \widetilde{y}(x)\right) = 3 * x - y(x) + Y_B * (20 - 4 * x + 2 * y(x)) * (4 * x - 2 * q * \widetilde{y}(x) - 20)$. As x is fixed, we rewrite agent B's utility function as $U_B\left(y(x), \widetilde{y}(x)\right)$. $U_B\left(y(x), \widetilde{y}(x)\right)$ is continuous in $y(x)$ and $\widetilde{y}(x)$, and $U_B\left(\cdot, \widetilde{y}(x)\right)$ is quasi-concave in $\widetilde{y}(x)$. By choosing a $y(x) \in G\left(\widetilde{y}(x)\right) = [0, 3 * x]$ 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 at least one $y(x) \in [0, 3 * x]$ that maximizes $U_B(y(x), \widetilde{y}(x))$ 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 set of $y(x) \in [0, 3 * x]$ which maximize $U_B(y(x), \widetilde{y}(x))$ 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. agent B's initial beliefs about agent A's beliefs about agent B's actions for x are correct. We apply Kakutani's Fixed Point Theorem and conclude that at least one fixed point exists.

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

Due to Lemma 1', it remains to show that given agent B's pure optimal strategy agent A has an optimal (possibly randomized) strategy a that is correctly expected by initial beliefs, i.e. $a = \widetilde{a}$ with $\widetilde{a}$ as agent A's initial second order belief on a.

Take any optimal pure strategy of agent B $y(x)$ for all $x \in \{0, 5, 10, 15, 20\}$ which is correctly expected by agent A. Then, agent A's utility function is $U_A\left(a, y(\cdot), \widetilde{a}\right) = \pi_A\left(a, y(\cdot)\right) + Y_A * \kappa_A\left(a, y(\cdot)\right) * \lambda_A\left(y(\cdot), \widetilde{a}\right)$. Let us define $E\left(x\right)$ and $\widetilde{E}\left(x\right)$ as the mean of x resulting with strategy a and $\widetilde{a}$, respectively, and $E\left(y\left(x\right)\right)$ and $\widetilde{E}\left(y\left(x\right)\right)$ as the mean of $y\left(x\right)$ resulting with strategy a and $\widetilde{a}$, respectively. Then, $\pi_A\left(a, y(\cdot)\right) = 20 -$

$E(x) - q*E(y(x))$, $\kappa_A(a, y(\cdot)) = 3*E(x) - q*E(y(x)) - (20 - E(x) + q*E(y(x)))$, and $\lambda_A(y(\cdot), \widetilde{a}) = 20 - \widetilde{E}(x) + q*\widetilde{E}(y(x)) - \left(3*\widetilde{E}(x) - q*\widetilde{E}(y(x))\right)$. Hence, $U_A(a, y(\cdot), \widetilde{a}) = 20 - E(x) - q*E(y(x)) + Y_A*(4*E(x) - 2*q*E(y(x)) - 20)*\left(20 - 4*\widetilde{E}(x) + 2*q*\widetilde{E}(y(x))\right)$. As $y(\cdot)$ is fixed, we can rewrite agent A's utility function as $U_A(a, \widetilde{a})$. $U_A(a, \widetilde{a})$ is continuous in a and $\widetilde{a}$, $U_A(\cdot, \widetilde{a})$ is quasi-concave, and agent A's set of possibly randomized strategies 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 strategies $X^*(\widetilde{a})$ out of which each strategy 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, and 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] Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games and Economic Behavior 10, 122-142.

[3] 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.

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

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

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

[7] Charness, G., Levine, D. I., 2007. Intention and stochastic outcomes: An experimental study. The Economic Journal 117, 1051-1072.

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

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

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

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

[14] Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10, 171-178.

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

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

[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. Mimeo.

[19] McCabe, K. A., Rigdon, M. L., Smith, V. L., 2003. Positive reciprocity and intentions in trust games. Journal of Economic Behavior and Organization 52, 267-275.

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

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

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

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

[24] Trautmann, S. T., 2009. A tractable model of process fairness under risk. Journal of Economic Psychology 30, 803-813.