DEPARTMENT OF ECONOMICS

UNIVERSITY OF MILAN - BICOCCA

WORKING PAPER SERIES

How to be kind? Outcomes versus Intentions as Determinants of Fairness

Luca Stanca

No. 145 – July 2008

Dipartimento di Economia Politica
Università degli Studi di Milano - Bicocca
http://dipeco.economia.unimib.it

How to be kind? Outcomes versus Intentions as Determinants of Fairness*

Luca Stanca†

June 2008

Abstract

This paper presents an experimental analysis of the role of outcomes and intentions for fair behavior. We consider a symmetric version of the gift-exchange game in a 2x2 design with two treatment variables: intentionality (first mover's choice is either intentional or randomly determined) and outcome (first mover's choice is either costly or free, ie compensated by the experimenter). The four treatments differ with respect to the presence-absence of intentionality and cost for the first mover, whereas the outcome of the first mover's action for the second mover's payoff is kept constant across treatments. The results indicate that intentions do not matter for fair behavior, whereas outcomes do matter. In particular, the effect of outcomes is due to concerns for distributional fairness, whereas there is no evidence of an intention-based role for outcomes through signalling kindness.

Keywords: Reciprocity, Effort, Intentions, Laboratory Experiments.
JEL codes: D63, C78, C91.

*We thank Beatrice Petrovich and Tommaso Reggiani for excellent research assistance. We acknowledge financial support from Econometica.

†Corresponding author. Economics Department, University of Milan Bicocca. Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy. E-mail: luca.stanca@unimib.it

1 Introduction

Recent economic research has indicated that economic behavior in many cases motivated by concerns for fairness. This evidence has led to the development of theoretical models that incorporate fairness in agents' preferences (see e.g. Fehr and Gachter, 2000, Sobel, 2004, and Fehr and Schmidt, 2006, for recent surveys). Alternative theoretical approaches differ with respect to how fairness is defined. In particular, two main classes of models can be distinguished: models that focus on distributional concerns, and models that focus on intention-based reciprocity.[1]

In the distributional approach, fairness refers to the *distribution* of material payoffs. Economic agents are motivated not only by their own material gain, but also by how their payoff compares with that of other agents. Fehr and Schmidt (1999) assume that the utility of a subject depends on the difference between his own payoff and that of other subjects, so that agents have egalitarian preferences. Bolton and Ockenfels (2000) assume that the utility function of a subject depends on his own payoff relative to the average overall payoff, so that agents care about their own relative status. In these models, fairness-related preferences depend only on the final distribution of payoffs, so that agents are not concerned about how a given distribution has been obtained.

In the reciprocity approach, fairness refers to the *intentions* of other agents. Agents derive utility from rewarding kind actions and punishing unkind actions, even if this is costly in terms of material payoffs (e.g. Rabin, 1993; Dufwenberg and Kirchsteiger, 2004). Preferences depend on the perceived kindness of an action and, therefore, on the beliefs about other agents' intentions (why an agent has chosen a given action).[2] In these models, actions with identical outcomes may elicit different reciprocating responses depending on how they are interpreted. A key question for intention-based reciprocity models is therefore *how* agents evaluate the kindness of a particular

[1] In this paper we use the term reciprocity to refer to strong reciprocity, defined as the non-strategic conditional behavior to reward kind actions, and to punish unkind actions, even if this is costly for the reciprocating subject.

[2] Both distribution and intentions play a role in the models by Charness and Rabin (2002) and Falk and Fischbacher (2006). In particular, in the theory of reciprocity by Falk and Fishbacher (2006) the kindness of an action depends on both intentionality and the outcome of an action, where the latter is defined as the difference in the payoffs of the receiving and sending subjects.

action.

At the empirical level, one way of assessing the kindness of an action is to compare the action intentionally chosen with the alternative actions that could have been chosen, thus focusing on the strategy space of the first mover. Both intentionality, intended as free-will, and the set of alternative possibilities therefore may contribute to define the perceived kindness of an action. This implies two testable predictions. First, there should be no intention-based reciprocal behavior when the action of the first mover is not chosen intentionally, for example because it is the only available option or it is determined exogenously, by a disinterested third party or by chance. Second, the perceived kindness of an intentionally chosen action depends on the characteristics of the alternative actions that were available to the agent but were not chosen.

A first group of experimental studies has investigated the role of intention-based fairness by focusing on the first prediction, testing the relevance of first mover's intentionality (the so-called attribution hypothesis). A control treatment where the sender can intentionally choose what action to take among a set of alternatives (thus signalling her intentions) is compared with a treatment where the sender cannot choose, either because she does not have alternative options, as in McCabe et al. (2003), or because her choice is determined randomly, as in Blount (1995) and Falk et al. (2008). The evidence, however, is mixed, and different results are obtained for positive and negative reciprocity.[3] It is important to observe that the notion of intentions investigated in this literature refers to the attribution of first mover's intentionality (free will).

A second group of experimental studies investigates the role of fairness intentions by focusing on the second testable prediction, testing the relevance of the alternative actions available to the first mover for the perceived kindness of a chosen action. In these studies, the strategy space of the first

[3]Bolton et al. (1998) study both positive and negative reciprocal behaviour, finding that distributional preferences are sufficient to explain observed reciprocal actions, whereas intentions play a marginal role. Blount (1995) finds significant evidence of attribution-based behavior only for negative reciprocity (see also Offerman, 2002). Charness (2004) compares a standard gift-exchange game to a treatment where the wage is determined randomly, finding that the slope of the relationship between wage and effort is significantly higher when wages are chosen by the employer. This lends some support to the role of intentions for positive reciprocity, although most of the reciprocal action can be attributed to distribution. Falk et al. (2008) find that the attribution of fairness intentions has a large and significant impact on both positive and negative reciprocal behavior.

mover is manipulated in ways that are strategically irrelevant, but potentially relevant for assessing the fairness of intentions (e.g. Andreoni et al., 2002, Brandts and Solà, 2001, Falk et al., 2003). These studies generally indicate that the perceived fairness of intentions is sensitive to alternative strategy spaces. Bolton and Ockenfels (2005) provide evidence that both distributional factors (relative shares) and strategy spaces (available actions) matter for fairness behavior. Bolton et. al. (2005) study experimentally the influence of procedural fairness on the pattern of acceptance and resistance to different outcomes, finding that choice behavior is sensitive to procedural fairness. Overall, however, the evidence on the role of non-distributional factors for models of social preferences is not conclusive. In particular, what determines the perceived kindness of an action remains an open question.

In this paper we compare the role of outcomes and intentions for fair behavior by examining experimentally a symmetric version of the gift-exchange game. While most experimental research has concentrated on either outcomes or intentions, we propose an experimental design that allows to model jointly the role of outcomes and intentions. The advantage of this design is that it allows to disentangle the effect of outcomes through distributional fairness, as in Bolton and Ockenfels (2000) and Fehr and Schmidt (1999), from the effect of outcomes as a signal of kindness through intention-based reciprocity, as in the models by Charness and Rabin (2002) and Falk and Fischbacher (2006).

The results indicate that outcomes matter for fair behavior, whereas intentions do not matter. In particular, the effect of outcomes is largely due to distributional fairness, whereas the effect of outcomes through signalling kind intentions, as in the model by Falk and Fischbacher (2006), plays a marginal role. These findings are robust to alternative ways of measuring fair behavior (in terms of amounts returned by second movers or correlations between amounts sent and returned) and to alternative test procedures (between subjects or within subjects). Overall, the results provide an indication that the role of intentions for fair behavior may have been overemphasized.

The paper is structured as follows. Section 2 presents the theoretical framework for the analysis. Section 3 describes the experimental design. Section 4 presents the results. Section 5 concludes with a discussion of the main findings and the implications of the analysis.

2 The Model

Following Falk and Fishbacher (2006), we present a framework to model the effect of both outcomes and intentions on fair behavior. Consider agent i, who is the second-mover in a one-shot sequential interaction with agent j. The utility function of agent i is assumed to depend not only on material payoffs (π_i), but also on concerns for fairness, represented by a distribution component and a reciprocity component. The distribution component is expressed as the product of a distributional sensitivity parameter and a distribution measure. The reciprocity component is given by the product of a reciprocity parameter, a kindness term and a reciprocation term:

$$U_i(\pi_i, \pi_j) = \pi_i + \alpha_i \delta_i + \rho_i \phi_i \sigma_i \tag{1}$$

The parameter α_i represents the agent's sensitivity to distributional factors. The distribution parameter (δ_i) measures distributional fairness.[4] The reciprocity parameter (ρ_i) represents the agent's sensitivity to reciprocity. The kindness term (ϕ_i) measures how kind the agent perceives the action undertaken by the other agent. The reciprocation term (σ_i) measures the effect of the reciprocal action on the other agent's utility. Depending on the relative size of the parameters α_i and ρ_i, and on the specification of δ_i, ϕ_i, and σ_i, the distributional and intention-based reciprocity components may have a different relative weight in the agent's preferences.

Focusing on the reciprocity component, in Falk and Fishbacher (2006), the kindness term depends on both the outcome and the intention underlying the action one is responding to:

$$\phi_i = \Delta_i \vartheta_i \tag{2}$$

where the outcome term Δ_i is defined as the difference between the second mover's payoff and the first mover's payoff ($\pi_i - \pi_j$), and the intention factor ϑ_i is a coefficient between 0 and 1 that parametrizes the intentionality of the action, with $\vartheta_i = 1$ describing a fully intentional action and $\vartheta_i < 1$ an action not fully intentional.

[4] In Fehr and Schmidt (1999) the utility function depends negatively on the difference between the agent's payoff and the payoff of the other agent. It is decreasing in the absolute difference, but the rate of decrease is greater for unfavorable inequality than for favorable inequality. In Bolton and Ockenfels (2000) the utility function, strictly concave in the agent's share of total payoffs, depends negatively on the difference between the agent's payoff and the average payoff of other agents.

Note that, within this framework, the relevance of intentions can be assessed by changing exogenously the parameter ϑ, capturing the degree of intentionality. If intentions matter, fair behavior should be elicited differently depending on the value of ϑ. The relevance of distributional outcomes can be assessed by varying δ_i. Note, however, that distributional outcomes can also be used to interpret the intentions of the first mover. Therefore, a change in the distribution of outcomes is also reflected in Δ_i. In our experimental design we use both ϑ and π_j as treatment variables. This enables us not only to identify and compare the effects of outcomes and intentions on fairness, but also to disentangle the effects of outcomes through distributional fairness (δ) and through intention-based reciprocity (Δ).

3 The Experiment

3.1 Experimental design

We examine experimentally a symmetric version of the gift-exchange game (e.g. Fehr et al. 1993, Gachter and Falk, 2002) that consists of two stages. At the beginning of the game, both players (A and B) are given an endowment of 20 tokens. In the first stage, player A must choose the amount a (an integer between 0 and 20) she wants to send to player B; the amount sent is subtracted from the payoff of A, multiplied by 3 by the experimenter, and added to the payoff of B. In the second stage, player B must choose the amount (an integer between 0 and 20) she wants to send to player A; the amount sent is subtracted from the payoff of B, multiplied by 3 by the experimenter, and added to the payoff of A. Total payoffs are therefore $20 - a + 3b$ for player A and $20 - b + 3a$ for player B. For each player the minimum and maximum potential payoffs are 0 and 80 tokens, respectively.[5]

In stage 2, when players B have to make their choice, we apply the strategy method (henceforth SM): player B has to provide a response for each feasible action of player A, *before* being informed of the actual choice of A. This allows us to study the responses to each possible action of A and therefore, on the basis of responses to different actions of A, to distinguish between unconditional altruism and conditional altruism (positive reciprocity) in the

[5]Note that symmetry in the endowments eliminates the confounding effects of distributional aspects, that may arise for example in a trust game: since both players have the same endowment, inequality aversion cannot determine A's decision.

strategies of B players.

We consider four treatments in a 2x2 between-within design. The first treatment, used as a benchmark, is a standard direct reciprocity setting (DIR), in which first movers' action is intentional and costly. To assess the role of intentions, we consider a second treatment where it is common knowledge for all subjects that A's choice is determined randomly by a computer. In this treatment (NO-I) A's action is therefore costly but not intentional. To assess the role of outcomes, we consider a third treatment where it is common knowledge for all subjects that first movers will be compensated by the experimenter for the amount sent, so that they will not bear any cost. In this treatment (NO-C) first movers' action is therefore intentional but not costly. We also consider a fourth treatment, where it is common knowledge that first movers' choice is determined randomly *and* they will be compensated by the experimenter for the amount sent. In this treatment (NO-CI) first movers' action is therefore neither intentional nor costly.

The four treatments therefore differ with respect to the presence-absence of either intentions or costs for the first mover, whereas the outcome of the first mover's action for the second mover's payoff is kept constant across treatments. The role of intentions can be assessed by comparing treatments 1 and 2 or treatments 3 and 4. As for the role of outcomes on fair behaviour, by comparing treatments 1 and 3 we can assess the effect through both distributional fairness and intention-based reciprocity (outcomes signal kind intentions). By comparing treatments 2 and 4 we can assess the pure effect of outcomes through distributional fairness.

3.2 Hypotheses

The experiment is designed to address the following hypotheses:

1. Intentions matter for fair behavior: an action is perceived to be kind depending on whether it is intentionally chosen by the first mover.

2. Outcomes matter for fair behavior: agents care for distributional fairness, as they prefer more equitable outcomes irrespective of intentions.

Our operational definition of fair behavior is based on two indicators. First, the amount returned by B players, for any given amount received. Second, the correlation coefficient between the amounts sent by A and B.

We focus on the Spearman coefficient, than Pearson correlations, so as to avoid restricting the attention to linear dependence.[6] Nevertheless, in order to enable a comparison of the two indicators, in presenting the results we will also report Pearson correlation coefficients.

Defining μ_i as the mean amount returned by B players or, alternatively, as the correlation between amounts sent and returned for treatment i, we test the following hypotheses:

1a. $H_0 : \mu_1 = \mu_2$ vs $H_a : \mu_1 > \mu_2$ (intentions matter)

1b. $H_0 : \mu_3 = \mu_4$ vs $H_a : \mu_3 > \mu_4$ (intentions matter in the absence of costs)

2a. $H_0 : \mu_1 = \mu_3$ vs $H_a : \mu_1 > \mu_3$ (outcomes matter)

2b. $H_0 : \mu_2 = \mu_4$ vs $H_a : \mu_2 > \mu_4$ (outcomes matter in the absence of intentions)

3. $H_0 : \mu_2 = \mu_3$ vs $H_a : \mu_2 \neq \mu_2$ (outcomes vs intentions)

Note that the relevance of distributional outcomes is assessed by varying the cost borne by the first mover. However, distributional outcomes can also be used to interpret the intentions of the first mover. Therefore, a change in the distribution of outcomes can also be reflected in stronger perceived kindness of an action. By comparing treatments 1 and 3 we can assess the effect of outcomes on fair behaviour through both distributional fairness and intention-based reciprocity (outcomes as a signal of kind intentions). By comparing treatments 2 and 4 we can assess the pure effect of outcomes through distributional fairness.

3.3 Procedures

We run four sessions, with 24 subjects participating in each session, for a total of 96 subjects. In each of the four sessions, subjects were randomly assigned to a computer terminal at their arrival. In order to ensure public knowledge, instructions were distributed and read aloud (see Appendix

[6]Note that if, for example, reciprocity was motivated only by inequality aversion, this could imply a linear relationship between b and a, so that Pearson correlations could be used as an appropriate indicator of reciprocating behaviour. However, if reciprocity is motivated also (or only) by the fairness of intentions, the relationship between b and a is not necessarily linear, so that the Spearman correlation coefficient is more appropriate.

1). Sample questions were distributed to ensure understanding of the experimental procedures. Answers were privately checked and, if necessary, explained to the subjects, and the experiment did not start until all subjects had answered all questions correctly.

In each session each subject played one phase for each of the four treatments, in the role of either A or B, for a total of 8 phases.[7] In each phase, subjects were assigned to their role as player A or B. A stranger matching mechanism was used to avoid strategic behavior. In addition, there was no information feedback after each phase, to avoid cross-subject dependence. Subjects therefore played 8 independent one-shot games. Only at the end of the 8 phases, one phase was selected randomly to determine payments and each subject was informed of her payoff in tokens and in euros.

The experiment was conducted in the Experimental Economics Laboratory of the University of Milan Bicocca in November 2008. Participants were undergraduate students of Economics recruited by e-mail using a list of voluntary potential candidates. Sessions lasted approximately 55 minutes. No show-up fee was paid and the exchange rate was 3 tokens = 1 euro. The average payment was 11.6 euros. The experiment was run using the experimental software z-Tree (Fischbacher, 2007).

4 Results

Table 1 and figures 1-2 report mean and median amount sent and returned for each treatment. The average amount sent by A players is 6.4 tokens in the DIR-treatment and 15.5 in the NO-C treatment, as opposed to 10 euros in treatments NO-I and NO-CI, where it is drawn randomly from a uniform distribution between 0 and 21. Average responses by players B are relatively similar in the DIR and NO-I treatments (5.23 and 5.53, respectively). They are also similar, but smaller, in the NO-C and NO-CI treatments (4.07 and 3.8, respectively). A similar pattern applies to returned amounts by individual input (from 0 to 21), although the differences between the treatments are more marked as the amount sent by the first mover rises. This pattern is shown clearly in figures 3 and 4, where mean and median amounts returned by second movers are displayed for each possible amount received. Both fig-

[7] The sequence of treatments was 1-2-3-4 in session 1, 2-1-4-3 in session 2, 3-4-1-2 in session 3, and 4-3-2-1 in session 4. The same sequence was used for each player in role A and B.

ures indicate that the profiles for the four treatments are relatively similar up to a received amount of about 5 tokens, whereas the diverge thereafter, with the averages for the two treatments in which the first movers are not compensated (DIR and NO-I) remaining well above the two treatments where the first movers are compensated (NO-C and NO-CI).

Tables 2 and 3 report test results for the hypothesis that median amounts returned by second movers are the same between pairs of treatments, both overall and by individual amount sent by first movers. In particular, table 2 reports results of rank-sum tests carried out between subjects, while table 3 reports results of sign-rank tests carried out within subjects. In both cases, tests are based on 96 independent observations.

The results for the tests between subjects (table 2) indicate that intentions do not matter: median returned amounts are not statistically different when first movers' intentionality is removed, both in the presence of costs for the first mover (comparing treatments DIR and NO-I, column 1) and in the absence of costs for the first mover (comparing treatments NO-C and NO-CI, column 6). On the contrary, outcomes do matter: differences in median returned amounts are strongly statistically significant when first movers are compensated, both in the presence of intentions for the first mover (comparing treatments DIR and NO-C, column 2) and in the absence intentions for the first mover (comparing treatments NO-I and NO-CI, column 5). Indeed, this latter result indicates that the concern for distributional fairness alone is sufficient to explain the role of outcomes. These results apply both to results over all possible amounts received (rows 1-2), and to results for individual amounts received, with the exception of low input levels (below 5 tokens).

The results for the tests within subjects (table 3) are more clear-cut in providing the same indications: intentions do not matter, whereas outcomes matter. The reason why outcomes matter is that agents appear to be concerned by distributional fairness, whereas outcomes are not used to interpret the perceived kindness of an action.

> **Result 1**: Intentions do not matter for fair behavior: amounts returned by second movers are not affected by whether first movers actions are intentional.

> **Result 2**: Outcomes matter for fair behavior: amounts returned by second movers are significantly affected by whether first movers actions are costly.

Table 4 reports, for each treatment, average correlation coefficients between amounts sent and returned, with the corresponding p-values for the null hypothesis of zero correlation. Averages are calculated over 96 individual correlation coefficients based on 21 observations. Spearman correlation is highest in DIR (0.4), and gradually diminishing in NO-I (0.36), NO-C (0.31) and NO-CI (0.27). A qualitatively similar pattern applies to Pearson correlation coefficients. Tables 5 reports results of rank-sum tests carried out between subjects for the hypothesis that median individual correlations are the same between pairs of treatments. The results indicate that differences in median correlations are not statistically significant when intentions are manipulated (DIR vs NO-I and NO-C vs NO-CI), whereas they are marginally significant when first movers' costs are varied (DIR vs NO-C and NO-I vs NO-CI). The results for sign-rank tests carried out within subjects, reported in table 6, are instead much more clear-cut: outcomes matter, whereas intentions do not matter.

> **Result 3**: Intentions do not matter for fair behavior: correlations between amounts sent by first movers and returned by second movers are not affected by whether first movers actions are intentional.

> **Result 4**: Outcomes matter for fair behavior: correlations between amounts sent by first movers and returned by second movers are significantly affected by whether first movers actions are costly.

5 Conclusions

This paper compared the role of distributional outcomes and intentions for fair behavior. We proposed a design that allows not only to identify and compare the effects of outcomes and intentions on fair behavior, but also to disentangle the effect of outcomes through distributional fairness, as in Bolton and Ockenfels (2000) and Fehr and Schmidt (1999), from the effect of outcomes as a signal of kindness through intention-based reciprocity, as in the models by Charness and Rabin (2002) and Falk and Fischbacher (2006). We proposed a new experimental design with two treatment variables: first movers' intentionality and cost. The analysis was based on two alternative

ways of measuring fair behavior (amounts returned by second movers or correlations between amounts sent and returned) and test procedures between subjects and within subjects.

The experimental analysis provided two main results. First, intentions do not matter. This finding is at odds with the results in many related experimental investigations (e.g. Nelson, 2002, Sutter, 2007). Indeed, second movers tend to return higher amounts in the treatments where intentions are absent. One possible interpretation of this seemingly counter-intuitive result is that intention-based reciprocity also depends on the motivation driving an action (see Stanca et al., 2008): an action driven by a strategic motivation can be perceived as less kind than an action not drive by strategic motives. Second, outcomes do matter for fair behavior. In particular, the effect of outcomes is largely due to concerns for distributional fairness, whereas the effect of outcomes through signalling kind intentions, as in the model by Falk and Fischbacher (2006), plays a marginal role. Overall, the results provide an indication that the role of intentions for fair behavior may have been overemphasized.

6 Appendix: Instructions

This appendix reports the instructions distributed on paper to the subjects. Paragraph headings indicate in brackets if the given subsection is common to both treatments or is specific to the relevant treatment.

Instructions *[common to all treatments]*

- Welcome and thanks for participating in this experiment.
- During the experiment you are not allowed to talk or communicate in any way with other participants. If at any time you have any questions raise your hand and one of the assistants will come to you to answer it.
- By following the instructions carefully you can earn an amount of money that will depend on your choices and the choices of other participants.
- At the end of the experiment the tokens that you have earned will be converted in euros at the exchange rate 2 tokens = 1 euro. The resulting amount will be paid to you in cash.

General rules *[common to all treatments]*

- There are 24 subjects participating in this experiment.
- The experiment takes place in 8 independent phases. Instructions for each phases will appear on the screen.
- In each phase 12 couples of two participants will be formed randomly and anonymously, so that in each phase you will interact with a different subject.
- Within each couple, the two subjects will be randomly assigned two different roles: A and B.
- Therefore, in each phase each subject will interact exclusively with the other subject in her pair, without knowing her/her identity, with the role (A or B) assigned with equal probability.

- The choices that you and the other subject will make in each and the corresponding outcomes will be communicated at the end of the experiment.
- At the end of the experiment only one of the 8 phases will be selected randomly and earnings for each participant will be determined on the basis of the selected phase.

How players interact

- Both A and B will receive an endowment of 20 tokens each.
- Player A will have to decide how many tokens (between 0 and 20) to send to player B. *[DIR only]*
- A number of tokens of A between 0 and 20 will be sent to player B. The number of tokens sent will be determined randomly by the computer, so that player A will make no choice. *[NO-I, NO-CI only]*
- Player A will fully be reimbursed by the experimenter for the amount sent to player B, so that player A will bear no cost. *[NO-C , NO-CI only]*
- We will triple the amount sent, so that B will receive 3 tokens for each token sent by A.
- Therefore:
  - A will obtain 20 tokens minus the tokens sent to B; *[DIR, NO-I]*
  - A will obtain 20 tokens; *[NO-C , NO-CI]*
  - B will obtain 20 tokens plus 3 times the tokens sent by A.
- Subject B will have to decide how many tokens (between 0 and 20) to send to player A.
- We will triple the amount sent, so that A will receive 3 tokens for each token sent by B.
- Therefore, in total:

  - A will obtain 20 tokens minus the tokens sent to B in phase 1 plus 3 times the tokens sent by B in phase 2.
  - A will obtain 20 tokens plus 3 times the tokens sent by B in phase 2. *[NO-C , NO-CI]*
  - B will obtain 20 tokens plus 3 times the tokens sent by A in phase 1 minus the tokens sent to A in phase 2.
  - B will obtain 20 tokens plus 3 times the tokens of A sent by the computer in phase 1 minus the tokens sent to A in phase 2. *[NO-I , NO-CI]*

- The choice of how many tokens B wants to send to A will be made as follows: *before being informed of how many tokens A sent to B in phase 1*, B has to decide how many tokens she wants to send to A for each of the possible amounts that A could have sent to him $(0, 1, ..., 20$ tokens). Since there are 21 possible cases, B has to make 21 choices.

- After B players have made their choice, only the one corresponding to the actual decision of A will be used to determine the earnings.

- The phase will end and total earnings for each subject for this phase will be determined as the sum of the earnings obtained in phase 1 and in phase 2.

7 References

Andreoni, J., Brown, P., Vesterlund, L., 2002. What Makes an Allocation Fair? Some Experimental Evidence. Games and Economic Behavior 40, 1-24.

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.

Bolton, G.E., Brandts, J. and Ockenfels, A., 1998. Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Game. Experimental Economics 1, 207-220.

Bolton, G.E., Ockenfels, A., 2000. ERC - A Theory of Equity, Reciprocity and Competition. American Economic Review 90, 166–193.

Bolton, G.E., Ockenfels, A., 2005. A Stress Test of Fairness Measures in Models of Social Utility. Economic Theory 25, 957-982.

Bolton, G.E, Brandts, J., Ockenfels, A., 2005. Fair Procedures: Evidence from Games Involving Lotteries. Economic Journal 115, 1054-1076.

Brandts, J., Solà, C., 2001. Reference Points and Negative Reciprocity in Simple Sequential Games. Games and Economic Behavior 36, 138-157.

Charness, G., 2004. Attribution and Reciprocity in a Simulated Labor Market: An Experimental Investigation. Journal of Labour Economics 22, 665-688.

Charness, G., Rabin, M., 2002. Understanding Social Preferences with Simple Tests. Quarterly Journal of Economics, 817-869.

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

Falk, A., Fehr, E., Fischbacher, U., 2008. Testing Theories of Fairness - Intentions Matter. Games and Economic Behavior 62, 287-303.

Falk, A., Fehr, E., Fischbacher, U., 2003. On the Nature of Fair Behaviour. Economic Inquiry 41, 20-26.

Falk, A., Fischbacher, U., 2006. A Theory of Reciprocity. Games and Economic Behavior 54, 293-315.

Fehr, E., Gächter, S., 2000. Fairness and Retaliation: The Economics of Reciprocity. Journal of Economic Perspectives 14, 159-181.

Fehr, E., Kirchsteiger, G., Riedl, A., 1993. Does Fairness Prevent Market Clearing? An Experimental Investigation. Quarterly Journal of Economics 108, 437-460.

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

Fehr, E., Schmidt, K.M., 2006. The Economics of Fairness, Reciprocity and Altruism: Experimental Evidence. In: Kolm, S.C., Ythier, J.M. (Eds). Handbook of the Economics of Giving, Altruism and Reciprocity 1, 615-691.

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

Fischbacher, U., Gachter, S., Fehr, E., 2001. Are People Conditionally Cooperative? Evidence from a Public Goods Experiment. Economics Letters 71, 397-404.

Frey, B., 1997. Not just for Money: An Economic Theory of Personal Motivation. Brookfield: Edward Elgar.

Gachter, S., Falk, A., 2002. Reputation and Reciprocity: Consequences for the Labour Relation. Scandinavian Journal of Economics 104, 1-26.

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.

Nelson, W., 2002. Equity or Intention: It Is the Thought That Counts. Journal of Economic Behavior and Organization, 48(4), 423-30.

Offerman, T., 2002. Hurting Hurts More that Helping Helps. European Economic Review 46, 1423-1437.

Rabin, M.,1993. Incorporating Fairness into Game Theory and Economics. American Economic Review 83, 1281–1302.

Sutter, M., 2007. Outcomes versus Intentions: On the Nature of Fair Behavior and Its Development with Age. Journal of Economic Psychology, vol. 28, no. 1, January 2007, 69-78.

Sobel J., 2005. Interdependent Preferences and Reciprocity. Journal of Economic Literature 43, 392-436.

Table 1: Mean and median amounts sent and returned, by treatment

	Means				Medians			
	DIR	NO-I	NO-C	NO-CI	DIR	NO-I	NO-C	NO-CI
Subject A	6.42	9.98	15.51	9.98	5.00	10.00	20.00	10.50
Subject B - SM all	5.23	5.53	4.07	3.87	1.00	1.00	0.00	0.00
Subject B - SM 0	1.84	2.30	1.76	1.99	0.00	0.00	0.00	0.00
Subject B - SM 1	2.51	2.98	2.11	2.26	0.00	0.00	0.00	0.00
Subject B - SM 2	2.60	3.25	2.56	2.49	0.00	0.50	0.00	0.00
Subject B - SM 3	3.00	3.73	2.73	2.71	0.00	1.00	0.00	0.00
Subject B - SM 4	3.43	4.05	3.22	3.10	0.00	1.00	0.00	0.00
Subject B - SM 5	3.91	4.45	3.36	3.23	1.00	1.00	0.00	0.00
Subject B - SM 6	4.34	4.99	3.80	3.48	1.00	2.00	0.00	0.00
Subject B - SM 7	4.66	5.33	3.97	3.59	1.50	3.00	0.00	0.00
Subject B - SM 8	4.97	5.74	3.98	3.66	2.00	3.00	0.00	0.00
Subject B - SM 9	5.30	5.91	4.05	3.81	2.50	3.00	0.00	0.00
Subject B - SM 10	5.52	6.11	3.98	4.06	2.50	3.00	0.00	0.00
Subject B - SM 11	5.38	6.19	4.27	3.86	2.00	3.00	0.00	0.00
Subject B - SM 12	5.76	6.28	4.29	4.05	2.00	3.00	0.50	0.00
Subject B - SM 13	6.14	6.35	4.92	4.52	3.00	4.00	0.50	0.00
Subject B - SM 14	6.49	6.34	5.00	4.54	3.00	4.00	0.50	0.00
Subject B - SM 15	6.77	6.68	5.08	4.72	2.50	5.00	0.00	0.00
Subject B - SM 16	7.05	6.55	4.89	4.72	4.00	2.50	0.00	0.00
Subject B - SM 17	7.24	6.98	4.92	4.94	3.00	5.00	0.50	0.00
Subject B - SM 18	7.52	6.99	5.14	5.08	4.00	3.50	1.00	0.00
Subject B - SM 19	7.26	7.19	5.24	5.25	2.00	5.00	1.00	0.50
Subject B - SM 20	8.09	7.67	6.24	5.22	4.50	5.50	1.50	0.00

Note: Means and median amounts are calculated over 96 independent observations.

Table 2: Tests of differences in amounts returned (between subjects)

	DIR NO-I	DIR NO-C	DIR NO-CI	NO-I NO-C	NO-I NO-CI	NO-C NO-CI
Overall	-1.43	5.34	6.64	6.81	8.06	1.35
(p-value)	0.15	0.00	0.00	0.00	0.00	0.18
By input (p-value)						
SM 0	0.66	0.90	0.85	0.75	0.81	0.94
SM 1	0.64	0.25	0.27	0.12	0.14	0.95
SM 2	0.34	0.46	0.28	0.11	0.06	0.79
SM 3	0.30	0.58	0.39	0.12	0.07	0.78
SM 4	0.46	0.60	0.40	0.24	0.13	0.76
SM 5	0.63	0.34	0.18	0.17	0.08	0.68
SM 6	0.51	0.33	0.16	0.12	0.05	0.69
SM 7	0.49	0.26	0.11	0.09	0.03	0.64
SM 8	0.49	0.20	0.10	0.06	0.02	0.71
SM 9	0.70	0.15	0.08	0.08	0.04	0.79
SM 10	0.70	0.09	0.10	0.04	0.05	0.99
SM 11	0.52	0.21	0.11	0.06	0.03	0.73
SM 12	0.73	0.14	0.08	0.07	0.04	0.75
SM 13	0.90	0.18	0.08	0.13	0.06	0.70
SM 14	0.95	0.15	0.07	0.15	0.07	0.69
SM 15	0.91	0.14	0.08	0.09	0.05	0.79
SM 16	0.70	0.05	0.04	0.12	0.09	0.90
SM 17	0.93	0.07	0.06	0.06	0.06	0.91
SM 18	0.67	0.06	0.03	0.12	0.07	0.77
SM 19	0.97	0.14	0.12	0.12	0.10	0.88
SM 20	0.67	0.23	0.03	0.39	0.06	0.27

Note: The table reports rank-sum tests between subjects, and corresponding p-values, for a two-sided null hypothesis of equal median between treatments. Results in rows 1-2 are based on the 21 possible amounts sent by A. All tests are based on 96 independent observations.

Table 3: Tests of differences in amount returned (within subjects)

	DIR NO-I	DIR NO-C	DIR NO-CI	NO-I NO-C	NO-I NO-CI	NO-C NO-CI
Overall	-1.43	5.34	6.64	6.81	8.06	1.35
(p-value)	0.15	0.00	0.00	0.00	0.00	0.18
By input (p-value)						
SM 0	0.66	0.90	0.85	0.75	0.81	0.94
SM 1	0.64	0.25	0.27	0.12	0.14	0.95
SM 2	0.34	0.46	0.28	0.11	0.06	0.79
SM 3	0.30	0.58	0.39	0.12	0.07	0.78
SM 4	0.46	0.60	0.40	0.24	0.13	0.76
SM 5	0.63	0.34	0.18	0.17	0.08	0.68
SM 6	0.51	0.33	0.16	0.12	0.05	0.69
SM 7	0.49	0.26	0.11	0.09	0.03	0.64
SM 8	0.49	0.20	0.10	0.06	0.02	0.71
SM 9	0.70	0.15	0.08	0.08	0.04	0.79
SM 10	0.70	0.09	0.10	0.04	0.05	0.99
SM 11	0.52	0.21	0.11	0.06	0.03	0.73
SM 12	0.73	0.14	0.08	0.07	0.04	0.75
SM 13	0.90	0.18	0.08	0.13	0.06	0.70
SM 14	0.95	0.15	0.07	0.15	0.07	0.69
SM 15	0.91	0.14	0.08	0.09	0.05	0.79
SM 16	0.70	0.05	0.04	0.12	0.09	0.90
SM 17	0.93	0.07	0.06	0.06	0.06	0.91
SM 18	0.67	0.06	0.03	0.12	0.07	0.77
SM 19	0.97	0.14	0.12	0.12	0.10	0.88
SM 20	0.67	0.23	0.03	0.39	0.06	0.27

Note: The table reports sign-rank tests within subjects, and corresponding p-values, for a two-sided null hypothesis of equal median between treatments. Results in rows 1-2 are based on the 21 possible amounts sent by A. All tests are based on 96 independent observations.

Table 4: Average correlations, within treatments

	DIR	NO-I	NO-C	NO-CI
Spearman correlation	0.40	0.36	0.31	0.27
P-value	0.00	0.00	0.00	0.00
Pearson correlation	0.39	0.35	0.31	0.26
P-value	0.00	0.00	0.00	0.00

Note: Averages are calculated for 96 individual correlation coefficients, each calculated over 21 observations.

Table 5: Tests of differences in correlations (between subjects)

	DIR NO-I	DIR NO-C	DIR NO-CI	NO-I NO-C	NO-I NO-CI	NO-C NO-CI
Spearman						
U-test	0.57	1.49	2.20	0.89	1.54	0.72
P-value	0.57	0.14	0.03	0.37	0.12	0.47
Pearson						
U-test	0.58	1.40	2.29	0.80	1.61	0.91
P-value	0.56	0.16	0.02	0.42	0.11	0.37

Note: The table reports rank-sum tests between subjects, and corresponding p-values, for a two-sided null hypothesis of equal median between treatments. All tests are based on 96 independent observations.

Table 6: Tests of differences in correlations (within subjects)

	DIR NO-I	DIR NO-C	DIR NO-CI	NO-I NO-C	NO-I NO-CI	NO-C NO-CI
Spearman						
U-test	1.00	2.37	3.15	1.38	2.82	1.31
P-value	0.31	0.02	0.00	0.17	0.00	0.19
Pearson						
U-test	0.09	1.84	3.10	1.12	2.38	0.63
P-value	0.93	0.07	0.00	0.26	0.02	0.53

Note: The table reports sign-rank tests within subjects, and corresponding p-values, for a two-sided null hypothesis of equal median between treatments. All tests are based on 96 independent observations.

Figure 1: Mean amount sent, by treatment

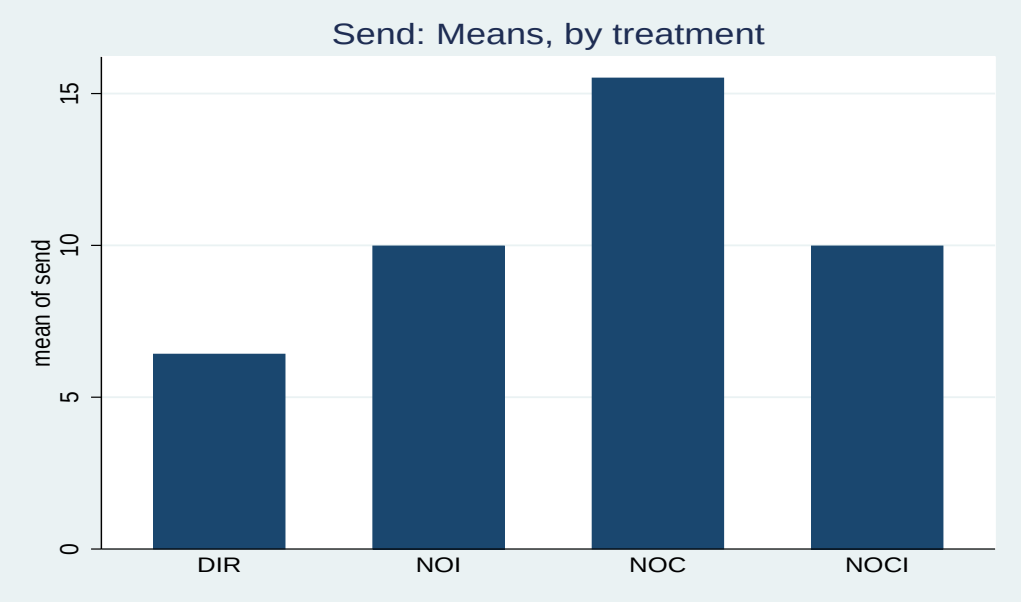

Figure 2: Mean amont returned, by treatment

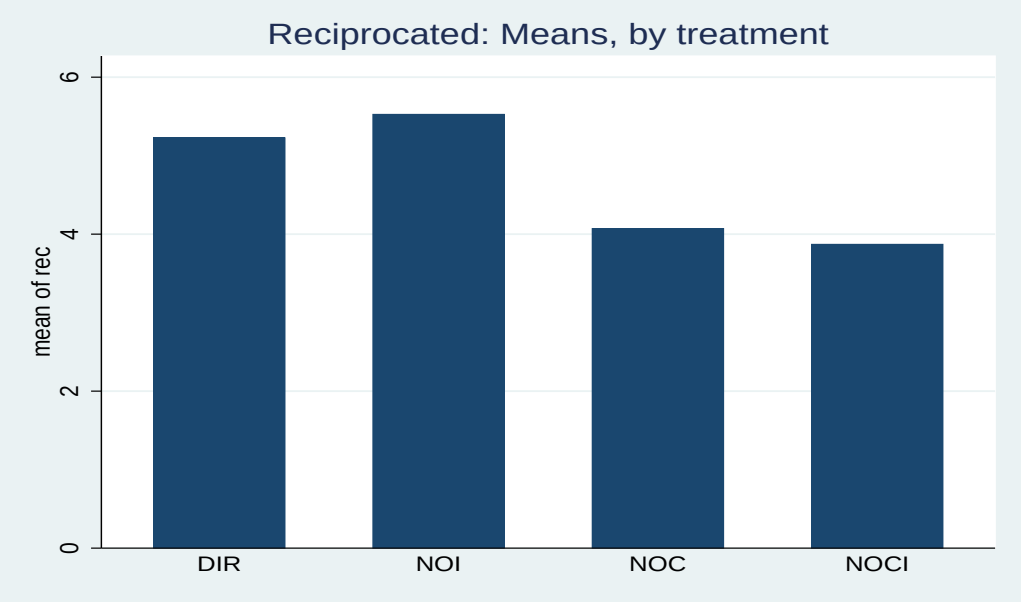

Figure 3: Mean amount returned, by treatment

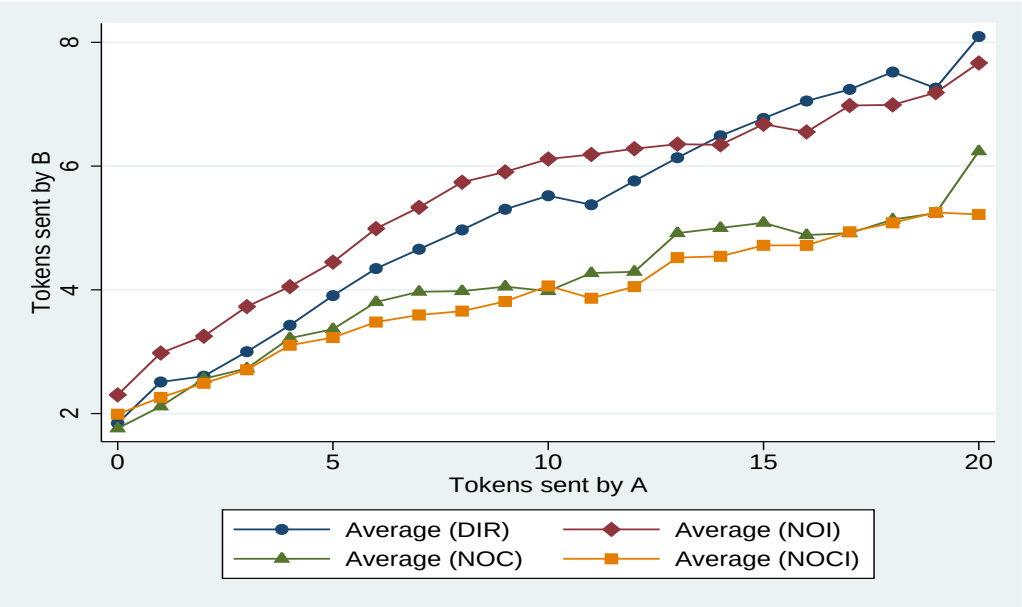

Figure 4: Median amount returned, by treatment