



DEPARTMENT OF ECONOMICS

UNIVERSITY OF MILAN - BICOCCA

WORKING PAPER SERIES

**The Effect of Motivations on Social Indirect  
Reciprocity: an Experimental Analysis**

Luca Stanca, Luigino Bruni, Marco Mantovani  
No. 169 – August 2009

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

# The Effect of Motivations on Social Indirect Reciprocity: an Experimental Analysis

Luca Stanca\*, Luigino Bruni<sup>†</sup>, Marco Mantovani<sup>‡</sup>

August 2009

## Abstract

This paper investigates the effects of motivations on the perceived kindness of an action within the context of strong social indirect reciprocity. We test experimentally the hypothesis that, for a given distributional outcome, an action is perceived by a third party to be less kind if it can be strategically motivated. The results do not support this hypothesis: social indirect reciprocity is indeed found to be significantly stronger when strategic motivations cannot be ruled out. We interpret these findings as an indication of the role played by team reasoning in explaining reciprocal behavior.

**Keywords:** Indirect Reciprocity, Motivations, Social Preferences, Laboratory Experiments.

**JEL codes:** D63, C78, C91.

---

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

<sup>†</sup>Economics Department, University of Milan Bicocca. Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy. E-mail: luigino.bruni@unimib.it

<sup>‡</sup>Graduate School in Social, Economic and Political Sciences, University of Milan. Email: marco.mantovani@unimi.it

# 1 Introduction

Indirect reciprocity is widely recognized in natural and social sciences, both theoretically and empirically, as an important mechanism in the evolution of cooperation (e.g. Axelrod, 1984, Nowak and Sigmund, 1998, Leimar and Hammerstein, 2001).<sup>1</sup> Indirect reciprocity is also becoming an increasingly widespread and relevant phenomenon in economic and social interactions, since iterated exchanges between partners tend to be replaced by one-shot interactions between strangers, as for instance in web-based auctions and other forms of e-commerce.<sup>2</sup>

A large number of experimental studies have investigated indirectly reciprocal behavior focusing on iterated interactions, where there can be a strategic incentive to reciprocate (e.g. Wedekind and Milinski, 2000, Engelmann and Fischbacher, 2003, Seinen and Schram, 2004, Bolton et al., 2005, Greiner and Levati, 2005).<sup>3</sup> Experiments based on repeated helping games (Nowak and Sigmund, 1998) generally provide evidence consistent with strategic reputation building: helping rates rise if agents know that their action will affect their own reputation.

Relatively less evidence is available on strong indirect reciprocity, defined as the conditional behavior to respond in kind to helpful and harmful actions by or towards a third party, when this is costly and provides neither present nor future net benefits for the reciprocator (e.g. Dufwenberg et al., 2001, Guth et al., 2001, Stanca, 2009). Despite the fact that one-shot interactions are becoming increasingly important, relative to repeated interactions, in many spheres of human life, the mechanisms underlying strong indirect reciprocity are still an open question and relatively little is known about its determinants.

In this paper, we focus on the behavioral relevance for strong social indirect reciprocity of the *motivation* driving the action an agent is responding to. (Rabin, 1993, Falk and Fischbacher, 2006).<sup>4</sup> This question is crucial in intention-based reciprocity models, but it is not properly addressed by the mere presence or absence of intentionality. We start by presenting a simple framework to model explicitly the effect of the nature of motivations on reciprocity. We then formulate and test the hypothesis that, for a given distributional outcome, an action is perceived to be less kind if it is strategically motivated (i.e., driven by the expectation of a higher future payoff) than if it is not. As a consequence, social indirect reciprocity is expected to be stronger

---

<sup>1</sup>See Nowak and Sigmund (2005) for a comprehensive review.

<sup>2</sup>Two different mechanisms of indirect reciprocity can be distinguished when an interaction between two agents also includes a third agent, not involved in the original exchange: social indirect reciprocity (IRS), where an action is reciprocated in kind at some cost *by* a third agent; generalized indirect reciprocity (IRG), where an action is reciprocated in kind at some cost *towards* a third agent.

<sup>3</sup>Such incentives are reputational in the case of social indirect reciprocity, or based on the expectation of cooperative behavior in the case of generalized indirect reciprocity.

<sup>4</sup>See Stanca et al. (2009) for an analysis of the effect of the nature of first-movers' motivations on second movers' reciprocating behavior in the case of direct reciprocity.

when strategic motivations by the first mover can be ruled out. We interpret this hypothesis as an interpersonal extension of motivational crowding-out, the reduction of effort in activities driven by intrinsic motivation following the introduction of an instrumental reward (Deci and Ryan, 2000).

In order to test this hypothesis, we design an experiment based on a gift-exchange game that allows us to manipulate the beliefs of the first mover about the strategy space of the second mover. The baseline gift-exchange game is in two phases: in phase 1 agent A decides whether to make a gift to agent B; in phase 2 agent C, informed of the action of A in phase 1, decides whether to make a gift to A. We compare two treatments varying the information sets of the players. In the first treatment, subjects are not aware of phase 2, so that giving in phase 1 cannot be strategic. In the second treatment, subjects are aware that there will be phase 2, so that giving in phase 1 can be driven by the expectation of a higher payoff in phase 2. The two treatments thus only differ with respect to the nature of the motivation of the first mover, whereas the distributional outcome of her action and her willingness to give are kept constant across treatments.

The results of the experimental analysis do not support the hypothesis of interpersonal motivational crowding out. Indeed, we find that reciprocity is *less* strong when strategic motivations can be ruled out. This result holds both for strategy profiles and for actual decisions. We provide an interpretation of the findings based on the role played by team reasoning (Sugden, 1993): the parties to a market transaction may understand it as a joint commitment to an activity whose aim is to pursue a common interest. When focusing on results at the individual level, three groups of agents emerge in the comparison of the two treatments: self-interested agents, intention-based reciprocators and common interest reciprocators. We find that it is the behavior of the last group that drives the different results across the two treatments.

The paper is structured as follows. Section 2 presents a framework to model the effect of the motivations of an agent on the reciprocal behavior of another agent. Section 3 describes the experimental design. Section 4 presents the results. Section 5 concludes with a discussion of alternative interpretations and implications of the findings.

## 2 The Effect of Motivations on Reciprocity

Two main classes of models incorporate fairness as a determinant of economic behavior (see e.g. Fehr and Gächter, 2000, Sobel, 2004, and Fehr and Schmidt, 2006, for recent surveys). In the outcome-based approach, fairness refers to the distribution of material payoffs (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), so that agents are not concerned about how a given distribution has been obtained. In the intention-based reciprocity approach, 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).

Within the intention-based approach, one of the key issues is what determines the perceived kindness of an action. In the following, building on Falk and Fishbacher (2006), we propose a framework to model the effect of the motivations driving the actions of an agent on the reciprocal behavior of another agent. 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 ( $\pi_i$ ), 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 expressed as the product of a reciprocity parameter, a kindness term and a reciprocation term:

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

The parameter  $\alpha_i$  represents the agent's sensitivity to distributional factors. The distribution measure ( $\delta_i$ ) measures distributional fairness.<sup>5</sup> The reciprocity parameter ( $\rho_i$ ) represents the agent's sensitivity to reciprocity. The kindness term ( $\phi_i$ ) measures how kind the agent perceives the action undertaken by the other agent. The reciprocation term ( $\sigma_i$ ) measures the effect of the reciprocal action on the other agent's utility. Depending on the relative size of the parameters  $\alpha_i$  and  $\rho_i$ , and on the specification of  $\delta_i$ ,  $\phi_i$ , and  $\sigma_i$ , the distributional and intention-based reciprocity components may have a different relative weight in the agent's preferences.

Focusing on the reciprocity component, the question we are considering is what determines  $\phi$ , the perceived kindness of an action. In Falk and Fishbacher (2006), the kindness term depends on both the outcome of the action and the underlying intention:

$$\phi_i = \Delta_i \vartheta_i \quad (2)$$

where the outcome term  $\Delta_i$  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  $\vartheta_i$  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.

In this paper we argue that the motivation driving an action is also relevant for its perceived kindness, so that the kindness term depends not only on the outcome of the action ( $\Delta$ ) and the intentionality the action ( $\vartheta$ ), but also on the type of motivation driving the action one is responding to (Gouldner, 1960). Perceived kindness therefore depends not only on whether an action

---

<sup>5</sup>In Fehr and Schmidt (1999) the utility function depends negatively on the difference between the agent's payoff and the payoffs of the other agent. 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.

produced a favorable outcome and on whether the action was intentionally chosen, but also on the reason why the action was chosen.

The expression for the kindness term in (2) should be extended as follows:

$$\phi_i = \Delta_i \vartheta_i \mu_i \quad (3)$$

where  $\mu$  is the *motivation factor*, a parameter between 0 and 1 that characterizes the type of motivation driving an action. We assume that an action is perceived to be less kind if it is strategically motivated ( $\mu_i < 1$ ) than if it is not ( $\mu_i = 1$ ).<sup>6</sup>

This implies that, for a given outcome ( $\Delta$ ) and intention ( $\vartheta$ ), the kindness term  $\phi$  is smaller in response to strategically motivated actions (S), than to non-strategically motivated actions (NS):

$$\phi_S < \phi_{NS} \quad (4)$$

or, alternatively,

$$\phi_S = (1 - \beta) \phi_{NS} \quad (5)$$

with  $\delta > 0$ . As a consequence, strategic motivations of the first mover result in weaker reciprocity of the second mover than non-strategic motivation ( $\beta > 0$ ).

This hypothesis can be interpreted as an interpersonal extension of the motivational crowding out hypothesis:<sup>7</sup> the intrinsic motivation of a strong reciprocator may be crowded out by the strategic motivation driving the kind action of the first mover. Existing studies of motivational crowding-out have generally investigated differences in the behavior of an agent motivated by either intrinsic or extrinsic rewards.<sup>8</sup> Our analysis examines whether the extrinsic motivation driving the action of an agent may have a crowding out effect on the intrinsically motivated indirectly reciprocal behavior of *another* agent.

In order to test this hypothesis, in the case of social indirect reciprocity, we designed an experiment based on a gift-exchange game with two treatments. In one treatment, the motivation of the first mover can only be perceived as non-strategic, while in the other treatment it can also be perceived as driven by the expectation of a higher future payoff.

---

<sup>6</sup>An action is strategically motivated if it is driven by the expectation of a higher future payoff.

<sup>7</sup>Motivational crowding out can be defined as the reduction of effort in activities carried out for intrinsic motivation when an instrumental reward, typically monetary, is introduced (e.g. Deci and Ryan, 2000).

<sup>8</sup>Frey (1997) examines the reduction of the effort of intrinsically motivated workers when extrinsic rewards are introduced. Gneezy and Rustichini (2000) explore the effects of motivational crowding-out in an experiment on fund-raising. At the theoretical level, Harvey (2005) explains motivational crowding out within a principal-agent model where utility is interpreted as overall satisfaction. The role of intrinsic motivation for optimal incentive contracts was examined by Kreps (1997) and Murdock (2002). See also the review in Lindenberg (2001).

### 3 Experimental Design

The experiment is based on a symmetric version of the gift-exchange game (e.g. Fehr et al. 1993, Gächter and Falk, 2002). This game has the advantage of making it easier for the reciprocating subject to interpret the nature of the other player's intentions, whose effect on reciprocity is the core of our analysis. The design is based on two treatments in a between-subjects design: full information (FI) and partial information (PI). We start by describing the details of the baseline game, then present the two treatments and the experimental procedures.

#### 3.1 Baseline Game

Subjects interact in groups of four with the following roles: A1, A2, B1 and B2. All subjects have the same initial endowment of 20 tokens and play the following two-stage gift-exchange game. In stage 1, A1 (A2) must choose the amount  $a_1$  ( $a_2$ ), an integer between 0 and 20, she wants to send to player B1 (B2). The amount sent is subtracted from the payoff of A1 (A2), multiplied by 3 by the experimenter, and added to the payoff of B1 (B2). In stage 2,  $B_i$  is informed of the amount sent by  $A_j$  in stage 1 and must choose the amount  $b_i$ , an integer between 0 and 20, she wants to send to player  $A_j$ ; the amount sent is subtracted from the payoff of  $B_i$ , multiplied by 3 by the experimenter, and added to the payoff of  $A_j$ .

At the end of stage 1, each subject is informed of her stage payoff in tokens. At the end of stage 2 each subject is informed of her stage payoff in tokens and of her total payoff in tokens and in euros. At the beginning of the game subjects are informed that no show-up fee will be paid, so that total earnings are determined only by total payoffs, converted with the exchange rate 2 tokens = 1 euro.

#### 3.2 Treatments

We compare two treatment of this game, where the treatment variable is the information set of players. In the FI-treatment, the whole game is perfectly known to players at the beginning of the game. That is, *before* playing stage 1 all subjects (A and B) are informed that there will be a stage 2. In the PI-treatment, instead, only *after* stage 1 has ended players (A and B) are informed that there will be a stage 2. In this case, stage 1 is played as if it were the whole game, and stage 2 is played as a surprise sub-game.<sup>9</sup>

While in the FI-treatment first movers are aware of the social indirect reciprocity mechanism when making their decision, they are not aware of it

---

<sup>9</sup>Note that subjects are simply given instructions about stage 1, without any explicit reference to the game ending thereafter, so that subjects are not in any way cheated by the announcement of stage 2. On the other hand, there is no reason why, when playing stage 1, subjects should expect stage 2 to follow. See the instructions in the appendix for details.

in the PI-treatment. In the FI-treatment, if player A thinks that reputation may be important in determining the choice of player B, she can expect net benefits from giving in stage 1. The same does not apply in the PI-treatment. Therefore, strategic incentives vary across treatments for the first mover, whereas they do not change for second movers (they are absent in both treatments). The two treatments therefore differ with respect to the motivations driving A's action. In the FI-treatment giving by  $A_i$  in stage 1 can be driven by pure altruism and/or by the expectation that  $B_j$  will reciprocate. In the PI-treatment, instead, giving cannot be driven by the expectation of a higher payoff in stage 2.

Since all players receive the same information, in stage 2 players B are fully aware of the motivation driving players A's actions in stage 1. Given that in stage 2 players B have to take exactly the same action as players A in stage 1, it is particularly simple for them to interpret the nature of A's motivations in each of the two treatments. Note that the equality of initial endowments is intended to minimize the relevance of distributional issues for A's choice. As for B's choice, any distributional motivation works in exactly the same way in the two treatments, so that differences in the behavior of B's cannot be accounted for with reference to distributive concerns. As all other conditions are kept fixed, any differences in the reciprocating behavior of players B can be interpreted as the effect of the differences in player A's motivation.

### 3.3 Hypotheses

If subjects are self-interested and rational, in both treatments second movers should give zero tokens for any number of tokens received, since they move at the terminal node and gift-giving is costly. In the FI-treatment, if it is common knowledge that all subjects are purely self-interested and rational, by backward induction in stage 1 the optimal choice of players A is to give zero. In the PI-treatment, the optimal choice of first movers is again to give zero, since they play as if they were at the terminal node of the game. Therefore, assuming that rationality is common knowledge, the subgame perfect equilibrium in both treatments is for all players to give zero.

On the other hand, if preferences are characterized by concerns for fairness, determined by either distributional outcomes or intention-based reciprocity, in both treatments the response of players  $B_i$  should depend positively on the amount sent by player  $A_j$ :

**Hypothesis 1 - Reciprocity:** in both treatments, the amount sent by  $B_i$  in stage 2 should be positively related to the amount sent by  $A_j$  in stage 1.

$$H_0: \rho_{FI} = 0 \text{ vs } H_1: \rho_{FI} > 0$$

$$H_0: \rho_{PI} = 0 \text{ vs } H_1: \rho_{PI} > 0$$

where  $\rho$  is the Spearman correlation coefficient between the amounts sent by subjects  $A_j$  and the corresponding choices of players  $B_i$ .<sup>10</sup>

Note that if some players are motivated by concerns for fairness, and this is common knowledge, the predictions for players A will differ in two treatments. In particular, players A should send more in the FI-treatment, since they might be motivated not only by pure altruism or concern for efficiency, but also by the expectation that indirectly reciprocal responses could increase their own payoff. This explains why, as explained below, the focus of our analysis is on the responses of players B in the strategy method, as it allows us to compare the two treatments, characterized by different motivations of first movers, while controlling for differences in the sending behavior of players A. We also analyze reciprocity in the actual responses of players B in order to provide a check of the robustness of our results.

Let us now turn to the key hypothesis to be tested. As outcomes cannot account for differences in the behavior of second movers across treatments, we consider the effect of different senders' motivations as a possible determinant of reciprocity. In particular, we expect social indirect reciprocity to be stronger in the PI-treatment, where extrinsic motivations can be ruled out, than in the FI-treatment.

**Hypothesis 2 - Effect of motivations on reciprocity:** The positive relationship between the amounts sent by B in stage 2 and by A in stage 1 should be stronger in the PI-treatment than in the FI-treatment.

$$H_0: \rho_{PI} = \rho_{FI} \text{ vs } H_1: \rho_{PI} > \rho_{FI}$$

### 3.4 Procedures

In stage 2 we applied a variant of the strategy method (Selten, 1967) to elicit subjects' preferences. This means that player  $B_i$  had to provide a response for each feasible action of player  $A_j$ , *before* being informed of her actual choice. We used the strategy method (henceforth SM) for two reasons. First, it allowed us to compare reciprocating behavior across treatments while keeping constant the actions of first movers, whose motivations are varied across treatments. Second, it allowed us to study conditional behavior at the individual level, thus making it possible to distinguish between unconditional altruism and positive reciprocity in the strategies of B players.<sup>11</sup>

After providing a response for each feasible action of player  $A_j$ , players  $B_i$  were informed of the actual action taken by  $A_j$  and had to choose a response (decision method, henceforth DM). Before players B made their

---

<sup>10</sup>We use Spearman correlations in order to avoid restricting our analysis to linear dependence. Pearson coefficients will be also reported to allow for the comparison of the two indicators.

<sup>11</sup>See e.g. Fischbacher et al. (2001) and Falk et al. (2008) for experimental analyses of reciprocity based on the strategy method.

choices with the two methods (SM and DM), all players were informed that payoffs would be determined on the basis of one of the two methods, to be selected randomly by publicly tossing a coin. After players B had made their decisions in both SM and DM, the method to determine the payoffs was selected on the basis of the outcome of the coin toss. This procedure based on responses by players B in both strategy and decision method allowed us to compare the consistency between the strategies of B players and their actual responses. It also allowed us to ensure that players  $A_j$  could choose their action in stage 1 knowing that in stage 2 players  $B_i$  would choose their action having been informed of the actual action taken by  $A_j$  in stage 1.

We run two sessions for each treatment, with 24 subjects participating in each session, for a total amount of 96 subjects. At the beginning of each session, subjects were randomly assigned a computer terminal and a role among  $A_1$ ,  $A_2$ ,  $B_1$  and  $B_2$ . Instructions for the game were distributed and read aloud (see Appendix). The experimental sessions were conducted in the Experimental Economics Laboratory of the University of Milan Bicocca. Participants were undergraduate students recruited by e-mail using a list of voluntary potential candidates. Sessions lasted approximately 45 minutes. The average payment was 13.5 euro. The experiment was run using the experimental software Z-tree (Fischbacher, 2007).

## 4 Results

Table 1 reports summary statistics by treatment and by player type. In the FI-treatment, the average amount sent is 7.54 tokens, while the median is 10 tokens. Things are rather different in the PI-treatment: mean and median amounts sent fall to 1.79 and 0.5 tokens, respectively. The differences between treatments in the amounts sent by first movers are statistically significant using a Mann-Whitney U-test. This finding suggests that in the FI-treatment first movers may be driven to give by the expected reciprocal behavior of second movers.<sup>12</sup>

The actual decisions of B-players are in line with the first movers actions, displaying a positive and a zero median amount sent, respectively, in the FI and PI treatments. Once we control for A's choice, using the strategy method, we find a positive median amount sent also in the PI-treatment. Although this is smaller than in the FI-treatment, the difference is not significant at the 5% level. Figures 2 and 3 display the distributions of the amounts sent by B players in the second stage.

Table 2 reports Spearman and Pearson correlation coefficients between first and second mover's choices by treatment, both for strategies and actual decisions. Figure 4 displays the corresponding histograms. Positive and

---

<sup>12</sup>Note, however, that other studies on indirect reciprocity show that first movers give positive amounts even in the absence of strategic incentives, as in the case of Generalized Indirect Reciprocity, where the recipient of the first mover is asked to give to a third player (Dufwenberg et al., 2001, Stanca, 2009). This kind of setting rules out strategic incentives for all of the roles involved.

significant coefficients are found for both treatments in strategy method. The non-significant results for the PI-treatment in DM reflect the high concentration of A-players' choices around zero. Overall, we can reject the null hypothesis of no reciprocity. A visual description of this result is provided by Figures 5 and 6, where the mean and median amounts returned, respectively, are plotted against the amount sent by A.

**Result 1:** There is significant evidence of strong social indirect reciprocity in both FI and PI treatments.

Table 2 also provides qualitative information regarding our main hypothesis. The size of the correlation coefficients in the two treatments does not support the hypothesis of motivational crowding out. To the contrary, the difference between the treatments goes in the opposite direction, displaying weaker reciprocity in the PI-treatment. When strategic incentives for the first mover are ruled out, the reputation effect is less effective in driving reciprocal responses. This result, albeit purely descriptive, is at odds with hypothesis 2.

Table 3 displays results at individual level, by ranking pairs of subjects on the basis of correlations in each of the two treatments. The corresponding histograms are shown in Figure 4. These figures, together with the corresponding test results reported in Table 4, confirm at individual level that lower levels of reciprocity are displayed in the PI-treatment.

**Result 2:** There is no evidence of motivational crowding out for social indirect reciprocity, both at the aggregate and at the individual level. To the contrary, reciprocal behavior is less strong when strategic motivations of first movers can be ruled out.

A closer look at the individual data, however, suggests that this may not be the whole story. Both individual responses in strategy method (Figures 7 and 8) and the corresponding correlations (Table 3) show that the weaker reciprocity observed on aggregate in the PI-treatment reflects a polarization in individual reciprocating behavior: about a half of the subjects display close-to-zero correlations, while the other half keeps reciprocating with close-to-one correlation coefficients, whereas in the FI-treatment individual correlations increase more gradually while displaying no discontinuity.

A possible interpretation of this result is based on the identification of different subject types. In both treatments, some second movers appear to be self-interested rational decision makers. In the FI-treatment, this type accounts for about one third of the sample.<sup>13</sup> The rest of the players reciprocate to different degrees. In the PI-treatment, the percentage of zero-givers is around 50 per cent and several other players send positive amounts, but independently of A's behavior. The rest of the group keeps reciprocating in different degrees.

---

<sup>13</sup>This estimate is in line with the bulk of the relevant literature [e.g. Fehr et al., 2000]

**Result 3:** Relative to the FI-treatment, there is a higher share of non-reciprocators in the PI-treatment, consisting of both zero-givers and unconditional altruists.

We interpret this result by considering a population of heterogeneous agents: along with some rational players, there appear to coexist a fraction of *intention-based* strong reciprocators and a fraction of *team reasoning* reciprocators. Consider the behavior of second movers. The first group will give zero in both treatments, while the second will always reciprocate. Motivational crowding out was predicted for this group, but is not confirmed by the data: they act in the same way in both treatments. The third group is responsible for most of the change across treatments. While behaving according to a reciprocity rule in the FI-treatment, they switch to a sending pattern that is independent from first mover's actions in the PI-treatment.

A possible explanation of this behavior is that team reasoning is based on a context where people share a common interest and act according to it, conditional on others doing the same. These requirements are satisfied in the FI-treatment, but not in the PI-treatment: team reasoning does not apply to the PI-treatment because the first mover's action cannot be aimed to a common interest. As a consequence, it would not make sense for a common interest reciprocator to condition her own behavior on it. Therefore, this group of agents will not reciprocate in the PI-treatment, and they will send positive or zero amounts just as the A-players did: as pure gifts.

It should be observed that the same mechanism, based on team-reasoning, is consistent with the pattern of A-players' behavior in stage 1. As argued by Sugden (1993, 2000), first movers may send positive gifts in a context where everybody receives and gives at his turn, just because it is in the group's common interest, and thus in each subject's interest, to do so. In the PI-treatment, along with strategic incentives, we also removed this additional reason for positive giving. The fall in giving rates in the PI-treatment can be accounted for not only by *missing strategic incentives*, but also by a *missing common interest*.

## 5 Discussion and Conclusion

This paper investigated the relevance of the nature of the motivations driving sending behavior in eliciting indirect reciprocity by a third party, focusing on the case of social indirect reciprocity. Beyond its theoretical interest, social indirect reciprocity is the relevant model of interaction in several domains. In particular, we consider Corporate Social Responsibility (CSR) as a natural application of our analysis of the effect of motivations on indirect reciprocity: the investigation of whether the motivations of the sender affect reciprocators' responses is of key interest for the evaluation of the effectiveness of CSR policies.<sup>14</sup>

---

<sup>14</sup>Graafland and van de Ven (2006) find that, in a sample of 111 Dutch firms, intrinsic motives are more important in inducing involvement of the stakeholders in CSR activities.

Using a baseline gift-exchange game, we tested the hypothesis that, for a given distributional outcome, an action towards an agent is perceived by a third party to be less kind if it strategically motivated than if it is not. The results do not support this hypothesis: reciprocity is indeed significantly stronger when strategic motivations of first movers cannot be ruled out. Individual-level data suggest that different types of agents are present and that they are responding differently to the two treatments. Interestingly, they fit three important theoretical models of behavior: self-interest, intention-based reciprocity and team reasoning. Behavior consistent with team reasoning is the most important in explaining different patterns of reciprocity in the two treatments. These findings have important implications for our understanding of social preferences and, in particular, for theories of reciprocity.

Common to all theories of *social preferences* is the idea that an individual is willing to sacrifice his own material interests to achieve some ‘other-regarding’ objective, such as equality between his own payoff and those of others or rewarding others for kind behavior. One approach to the modelling of social preferences is to assume *inequity aversion*, as in the theories proposed by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000). The social orientation of inequity aversion is represented as a self-sacrificing motivation – as the willingness to *give up* material payoffs to generate outcomes in which one’s own position relative to others is less unfair.

Theories of inequity aversion do not take into account the role of *intentions*. They thus fail to recognize that, for instance, in a gift-exchange game the second mover’s motivation to send a positive amount may be a response to the intentions that are revealed by the first mover’s choice. Rabin’s (1993) theory of reciprocity provides a formal representation of mutually responsive intentions. The idea is that individuals are motivated by kindness towards people who are being kind to them, and by unkindness to those who are unkind to them. Our paper pointed at motivations as possible determinants of the perceived kindness. As in the theory of inequity aversion, these other-oriented motivations are modelled in terms of self-sacrifice: to be kind to another person is to be willing to sacrifice one’s material payoff to benefit her, and to be unkind to another person is to be willing to sacrifice one’s material payoff to harm her.<sup>15</sup>

An alternative interpretation of our experimental results is offered by *team reasoning* theory (Sugden 1993; Bruni and Sugden 2008). The core idea of team reasoning, or collective intentionality, is that the parties to a market transaction understand it as a joint commitment to an activity whose aim is to benefit them all, pursuing a common interest. Mutual trust is a

---

<sup>15</sup>As part of their extension of Rabin’s theory to sequential-move games, Dufwenberg and Kirchsteiger (2004) propose an amendment to Rabin’s definition of kindness. It might seem that what is needed is a composite model of social preferences that takes into account inequity aversion *and* Rabin reciprocity *and* trust responsiveness (Battigalli and Dufwenberg 2005, Pelligra 2005). Such a theory has been proposed by Falk and Fischbacher (2006), who describe it as ‘a theory of reciprocity’.

facet of that commitment. This approach allows us to understand reciprocity as the product of individual motivations that are positively oriented towards others without being self-sacrificing.

Team reasoning provides a theory in which an individual's social orientation is represented as a positive attitude towards mutually beneficial practices.<sup>16</sup> That is, each player understands his own action as a component of a joint action that is intended by the players together. Notice that, in this account, reciprocity involves the reciprocation of *good* intentions: only in this sense, there is a parallel with intention-based reciprocity. However, the intentions that are reciprocated are collective, not individual, and they are directed towards mutual benefit, not gift-giving.

Summing up, this paper argued that theoretical models of reciprocity should take explicitly into account the effect of perceived motivations on reciprocating behavior. In intention-based theories, reciprocity is an attitude of individuals who are concerned about how each is treating the other. The key concepts are kindness and unkindness, giving and taking. In theories based on team reasoning, instead, reciprocity is related to each individual being willing to do his part in a collective action directed at a common objective. The reciprocation is of individual efforts directed at a common goal. Future research on the determinants of reciprocity will have to take a broader perspective, allowing for both these components in agents' preferences.

---

<sup>16</sup>“The core idea is that, in a game which is perceived as providing opportunities for cooperation, the players reason as if they were engaged in a problem of collective choice, jointly choosing the *profile* of strategies that has the best consequences for them collectively... In contrast to conventional game theory, in which each player asks himself ‘What should I do, given what I can expect the others to do?’, the players ask themselves the unconditional question ‘What should we do?’” (Sugden 2009, pp. 12-13).

## 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 number of 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**

- There are 24 subjects participating in this experiment.
- In the experiment you will interact in a group of 4 subjects.
- You will not be informed about the identity of the other subjects in your group, nor will they be informed about yours.

### **How players interact** [*PI-treatment*]

- The four subjects in your group will be randomly assigned to one of four roles: A1, A2, B1, and B2.
- Each subject will receive an endowment of 20 tokens.
- A1 chooses how many tokens (between 0 and 20) to send to B1. At the same time, A2 chooses how many tokens (between 0 and 20) to send to B2.
- We will triple the amount sent, so that B1 and B2 will receive 3 times the tokens sent by A1 and A2, respectively.
- Therefore:
  - A1 will obtain 20 tokens minus the tokens sent to B1;
  - A2 will obtain 20 tokens minus the tokens sent to B2;

- B1 will obtain 20 tokens plus 3 times the tokens sent by A1
- B2 will obtain 20 tokens plus 3 times the tokens sent by A2.

**How players interact** [*FI-treatment*]

- The four subjects in your group will be randomly assigned to one of four roles: A1, A2, B1, and B2.
- Each subject will receive an endowment of 20 tokens.
- The experiment will take place in 2 phases.

PHASE 1

- A1 chooses how many tokens (between 0 and 20) to send to B1. At the same time, A2 chooses how many tokens (between 0 and 20) to send to B2.
- We will triple the amount sent, so that B1 and B2 will receive 3 times the tokens sent by A1 and A2, respectively.

PHASE 2

- (IRS) In phase 2 B1, informed of the amount that A2 sent to B2 in stage 1 and will choose how many tokens (between 0 and 20) to send to A2. At the same time B2, informed of the amount that A1 sent to B1 in stage 1, will choose how many tokens (between 0 and 20) to send to A1.
- We will triple the amounts sent, so that A2 will receive 3 times the tokens sent by B1 and A1 will receive 3 times the tokens sent by B2.
- The experiment will end. Overall earnings for each subject will be determined as the sum of the earnings obtained in the two phases:
  - A1: 20 tokens minus the tokens he sent to a B1 in phase 1 plus 3 times the tokens sent to him by B2 in phase 2.
  - A2: 20 tokens minus the tokens he sent to a B2 in phase 1 plus 3 times the tokens sent to him by B1 in phase 2.
  - B1: 20 tokens plus 3 times the tokens sent to him by A1 in phase 1 minus the tokens he sent to A2 in phase 2.
  - B2: 20 tokens plus 3 times the tokens sent to him by A2 in phase 1 minus the tokens he sent to A1 in phase 2.

**Instructions ON SCREEN - phase 1** [*common to both treatments*]

- You are A1. You have to decide how many tokens (between 0 and 20) to send to B1. We will triple the amount sent, so that A2 will receive 3 tokens for each token you send.

- In phase 2 B2 will be informed of the amount that you sent to B1 in stage 1 and will choose how many tokens (between 0 and 20) to send to you.

**Instructions ON SCREEN - phase 2** [*common to both treatments*]

- You are B1. You have to decide how many tokens (between 0 and 20) to send to A2. We will triple the amount sent, so that A2 will receive 3 tokens for each token you send.
- (IRS) Your choice will be made with two different methods:
  - Method 1: *before being informed of how many tokens A2 actually sent to B2 in phase 1*, you have to decide how many tokens to send to A2 for each of the possible amounts that he could have sent to B2 in phase 1 (0, 1, ..., 20 tokens). Since there are 21 possible cases, you have to make 21 choices.
  - Method 2: *after being informed of how many tokens A2 actually sent to B2 in phase 1*, you have to decide how many tokens you want to send to A2.
- After you have made your choice with both methods, earnings will be determined on the basis of one of the two methods, selected randomly.
  - If method 1 is selected, of the 21 choices that you had made, only the one corresponding to the actual decision of A1 will be used to determine the earnings.
  - If method 2 is selected, the single choice that you had made will be used to determine the earnings.

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

## References

- Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books, New York.
- Binmore, K.G. (1998). *Game Theory and the Social Contract, II: Just Playing*. MIT Press, Cambridge, MA.
- Bolton, G.E., Ockenfels, A., 2000. ERC: A Theory of equity, reciprocity and competition. *American Economic Review* 90, 166-193.
- Bolton, G.E., Katok, E., and Ockenfels, A. (2005) "Cooperation among Strangers with Limited Information about Reputation", *Journal of Public Economics*, 89, 2005, 1457-1468.
- Boyd, R., Richerson, P.J. (1989). The evolution of indirect reciprocity. *Social Networks*, 11, 213-236.
- Bruni, L., Sugden, R. (2008). Fraternity: why the market need not to be a morally free zone. *Economics & Philosophy*, 24, 35-64.
- Deci, E. L., Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.
- Deci E.L. and R.M. Ryan, (2000) "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions", *Contemporary Educational Psychology* 25, 54-67, pp. 56-60.
- Dufwenberg, M., Gneezy, U., Güth, W., van Damme, E. (2001). Direct versus Indirect Reciprocity: An Experiment. *Homo Oeconomicus*, 18, 19-30.
- Dufwenberg, M., Kirchsteiger, G. (2004). A Theory of Sequential Reciprocity. *Games and Economic Behavior*, 47, 268-298.
- Engelmann, D. and Fischbacher, U. (2002) Indirect Reciprocity and Strategic Reputation Building in an Experimental Helping Game. Working Paper n. 132, Institute for Empirical Research in Economics, University of Zurich.
- Falk, A., Fehr, E., Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior*, 62(1), 287-303.
- Falk, F., Fischbacher, U. (2006). A Theory of Reciprocity. *Games and Economic Behavior* 54(2), 293-315.
- Fehr, E., Gächter, S. (1998). Reciprocity and economics: the Economic Implications of Homo Reciprocans. *European Economic Review*, 42, 845-859.
- Fehr, E., Gächter, S. (2000). Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives*, Summer, 14, 159-81.

- Fehr, E., Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137-140.
- 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, 3, 817-868.
- Fehr, E., and Schmidt, K.M. (2006) The economics of Fairness, Reciprocity and altruism: Experimental evidence, in *Handbook of the Economics of Giving, Altruism and Reciprocity*, Eds: Varet, G., Ythier, Jean Mercier - Intriligator, Michael D. - Arrow, Kenneth J. - Kolm, S.-C.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10, 171-178.
- Fischbacher, U., Gächter, S., Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, Elsevier, vol. 71(3), 397-404.
- Frey, B., 1997. *Not just for Money: An Economic Theory of Personal Motivation*. Brookfield: Edward Elgar
- Gächter, S., Falk, A. (2002). Reputation and Reciprocity: Consequences for the Labour Relation. *Scandinavian Journal of Economics* 104(1), 1-26.
- Gintis, H. (2000). Strong Reciprocity and Human Sociality. *Journal of Theoretical Biology*, 206, 169-179.
- Gneezy, U., and A. Rustichini "Pay Enough or Don't Pay At All." *Quarterly Journal of Economics* August 2000, 791-810.
- Graafland, Johan and van de Ven, Bert (2006). Strategic and moral motivation for corporate social responsibility. *Journal of Corporate Citizenship* 22, 111-123.
- Greiner, B., Levati, M.V., (2005). Indirect reciprocity in cyclical networks: An experimental study. *Journal of Economic Psychology*, 26(5), 711-731.
- Guth, W., Königstein, M., Marchand, N., Nehring, K. (2001). Trust and Reciprocity in the Investment Game with Indirect Reward. *Homo Oeconomicus*, 18, 241-262.
- Kreps, D.M. (1997) Intrinsic Motivation and Extrinsic Incentives, *American Economic Review*, vol. 87, no. 2, pp. 359-64.

- Leimar, O., Hammerstein P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society Biological Sciences*, 268, 745-753.
- Lindenberg, S. (2001) Intrinsic Motivation in a New Light, *Kyklos*, vol. 54, n. 2-3, 2001, pp. 317-42.
- Murdock, K. (2002) Intrinsic Motivation and Optimal Incentive Contracts, *RAND Journal of Economics*, vol. 33, n. 4, pp. 650-71.
- Nowak, M., Sigmund, K. (1998). The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, 194, 561-574.
- Nowak, M.A., and Sigmund, K. (2005) Evolution of indirect reciprocity, *Nature*, Vol 437, n. 27, p. 1291-1298.
- Pelligra V. (2005), Under Trusting Eyes: the Responsive Nature of Trust. In Gui and Sugden (2005). *Economics and Social Interactions*. Cambridge University Press, Cambridge.
- Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review*. 83, 1281-1302.
- Seinen, I., Schram, A. (2006). Social Status and Group Norms: Indirect Reciprocity in a Repeated Helping Experiment. *European Economic Review*, 50(3), 581-602.
- Sobel J. (2004) Interdependent Preferences and Reciprocity, *Journal of Economic Literature*.
- Stanca, L. (2009). Measuring Indirect Reciprocity: Whose Back Do We Scratch?. *Journal of Economic Psychology*, 30(2), 190-202.
- Stanca, L., Bruni, L., Corazzini L. (2009). Testing Theories of Reciprocity: Do Motivations Matter?. *Journal of Economic Behavior and Organization*, 71(2), 233-245 .
- Sugden R. (1984). Reciprocity – The Supply of Public Goods through Voluntary Contributions, *Economic Journal*, 94, 772-787.
- Sugden, R. (1993). Thinking as a Team: toward an Explanation of Non-selfish behaviour, *Social Philosophy and Policy*, 10, 69-89.
- Sugden, R. (2009). Neither self-interest nor self-sacrifice: the fraternal morality of market relationships. Mimeo, University of East Anglia.
- Wedekind, C., Milinski, M. (2000). Cooperation through image scoring in humans, *Science*, 288, 850-852.

Table 1: Amount sent by treatment and significance of differences

|               | Mean FI | Mean PI | Med FI | Med PI | Ustat | Pval |
|---------------|---------|---------|--------|--------|-------|------|
| Player A      | 7.54    | 1.79    | 10.00  | 0.50   | 3.85  | 0.00 |
| Player B (SM) | 4.57    | 3.38    | 4.29   | 1.71   | 1.55  | 0.12 |
| Player B (DM) | 3.63    | 0.96    | 4.00   | 0.00   | 2.32  | 0.02 |

*Note:* FI = Full Information treatment. PI = Partial Information treatment.

Table 2: Reciprocity within treatments

|                      | FI (SM) | PI (SM) | FI (DM) | PI (DM) |
|----------------------|---------|---------|---------|---------|
| Spearman correlation | 0.49    | 0.27    | 0.56    | 0.20    |
| P-value              | 0.00    | 0.00    | 0.00    | 0.34    |
| Pearson correlation  | 0.50    | 0.32    | 0.53    | -0.07   |
| P-value              | 0.00    | 0.00    | 0.01    | 0.74    |

*Note:* FI = Full Information treatment. PI = Partial Information treatment. Columns 1 and 2 report correlations for individual strategies, columns 3 and 4 report correlations for individual decisions. All tests are based on 24 independent observations.

Table 3: Individual Spearman correlations, by treatment

|         | FI treatment | (P-value) | PI treatment | (P-value) |
|---------|--------------|-----------|--------------|-----------|
| Pair 1  | -0.06        | 0.81      | -0.69        | 0.00      |
| Pair 2  | 0.00         | 1.00      | -0.04        | 0.85      |
| Pair 3  | 0.37         | 0.10      | 0.00         | 1.00      |
| Pair 4  | 0.41         | 0.06      | 0.00         | 1.00      |
| Pair 5  | 0.43         | 0.05      | 0.00         | 1.00      |
| Pair 6  | 0.51         | 0.02      | 0.00         | 1.00      |
| Pair 7  | 0.75         | 0.00      | 0.00         | 1.00      |
| Pair 8  | 0.75         | 0.00      | 0.00         | 1.00      |
| Pair 9  | 0.75         | 0.00      | 0.00         | 1.00      |
| Pair 10 | 0.84         | 0.00      | 0.00         | 1.00      |
| Pair 11 | 0.87         | 0.00      | 0.00         | 1.00      |
| Pair 12 | 0.91         | 0.00      | 0.00         | 1.00      |
| Pair 13 | 0.93         | 0.00      | 0.46         | 0.04      |
| Pair 14 | 0.94         | 0.00      | 0.80         | 0.00      |
| Pair 15 | 0.97         | 0.00      | 0.91         | 0.00      |
| Pair 16 | 0.97         | 0.00      | 0.94         | 0.00      |
| Pair 17 | 0.97         | 0.00      | 0.96         | 0.00      |
| Pair 18 | 0.98         | 0.00      | 0.98         | 0.00      |
| Pair 19 | 0.98         | 0.00      | 0.99         | 0.00      |
| Pair 20 | 0.99         | 0.00      | 0.99         | 0.00      |
| Pair 21 | 1.00         | 0.00      | 0.99         | 0.00      |
| Pair 22 | 1.00         | 0.00      | 1.00         | 0.00      |
| Pair 23 | 1.00         | 0.00      | 1.00         | 0.00      |
| Pair 24 | 1.00         | 0.00      | 1.00         | 0.00      |

*Note:* FI = Full Information treatment. PI = Partial Information treatment. The figures reported are Spearman rank correlation coefficients, sorted by size, between the amount received and the amount sent by B players in strategy method.

Table 4: Test for differences in distributions of correlations

|                      | T-stat | P-val | U-stat | P-val | KS-stat | P-val |
|----------------------|--------|-------|--------|-------|---------|-------|
| Spearman correlation | 2.63   | 0.99  | 1.54   | 0.94  | 0.04    | 0.96  |
| Pearson correlation  | 2.76   | 1.00  | 1.74   | 0.96  | 0.08    | 0.85  |

*Note:* the table reports results of alternative tests of the null hypothesis that average correlation coefficients are the same in the Full Information treatment and the Partial Information treatment, versus the corresponding one-sided alternative.

Figure 1: Mean amount sent, by treatment

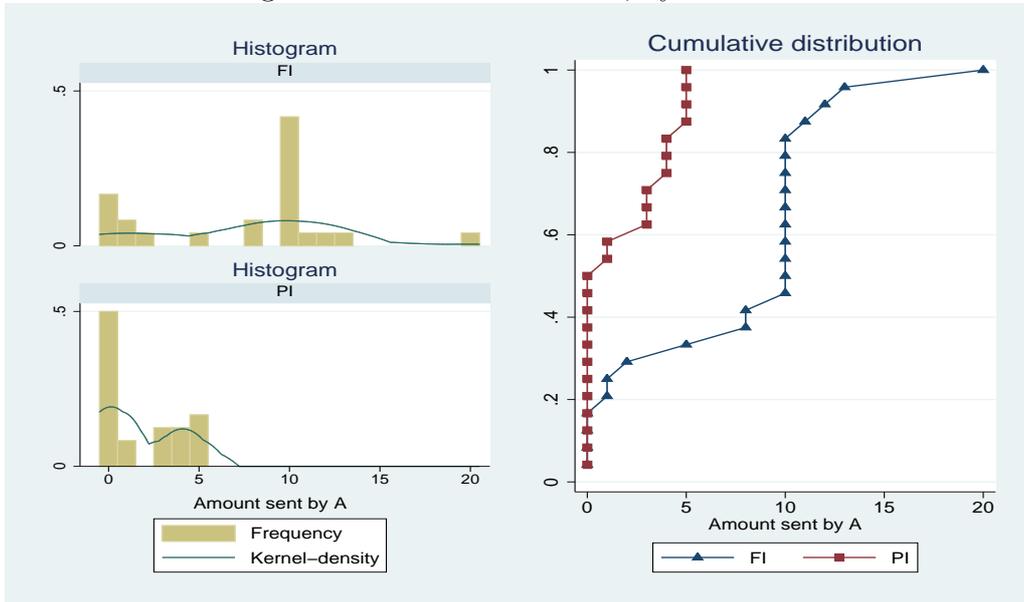


Figure 2: Mean amount returned, by treatment - Decision Method

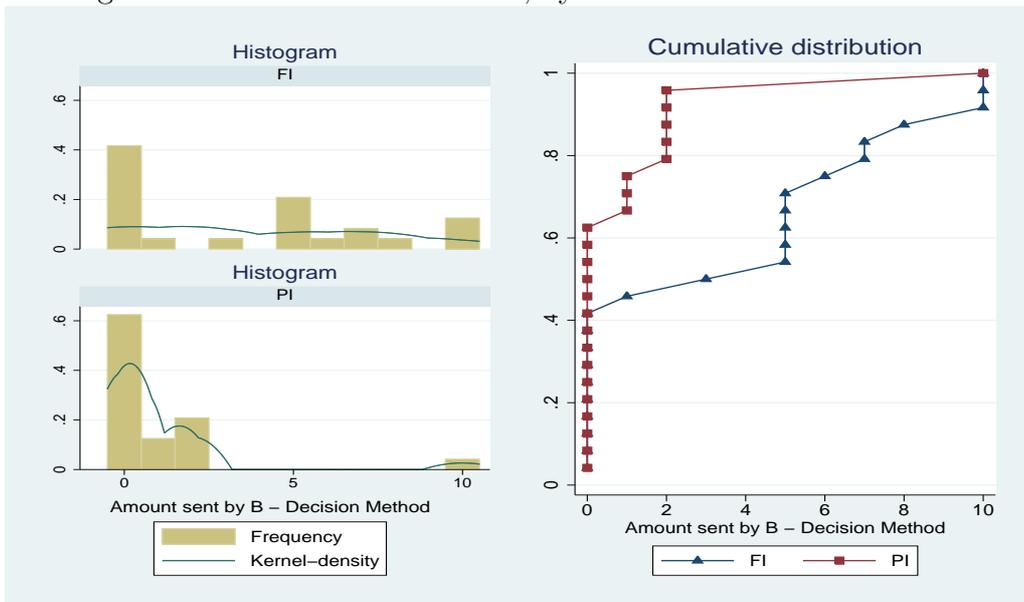


Figure 3: Mean amount returned, by treatment - Strategy Method

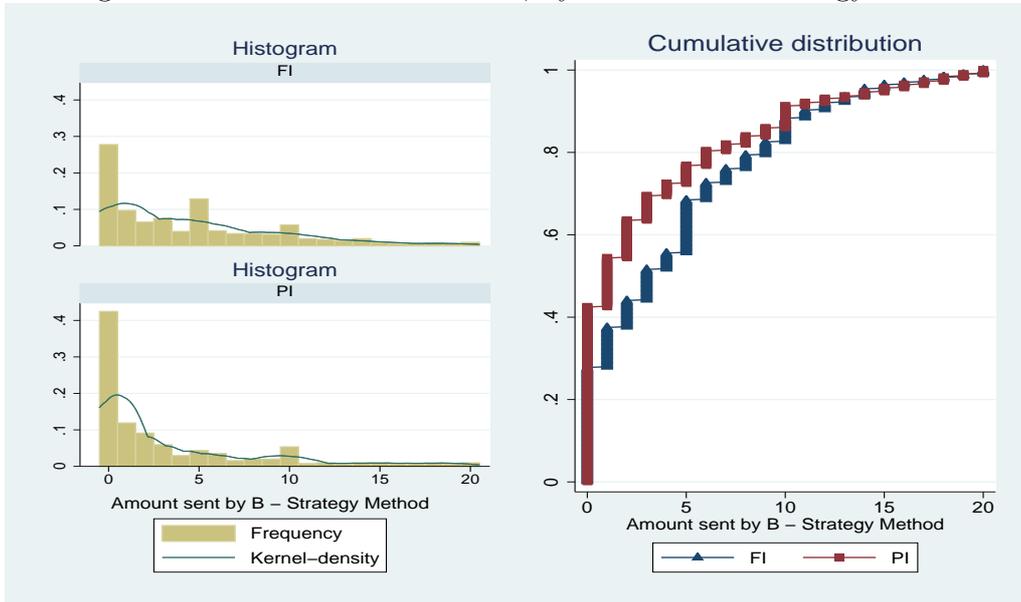


Figure 4: Distribution of Spearman correlations, by treatment

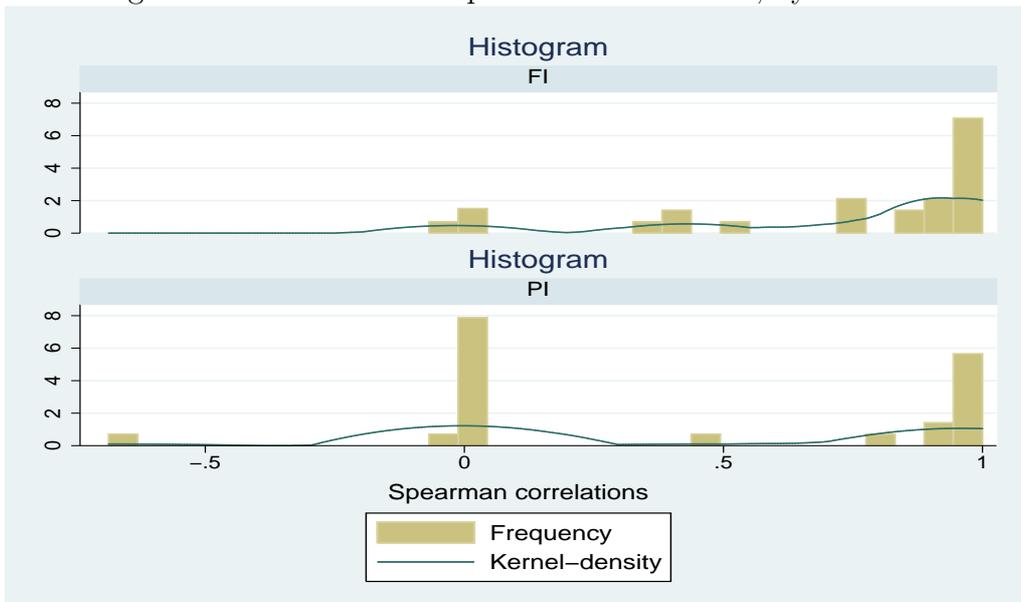


Figure 5: Mean amount returned, by treatment

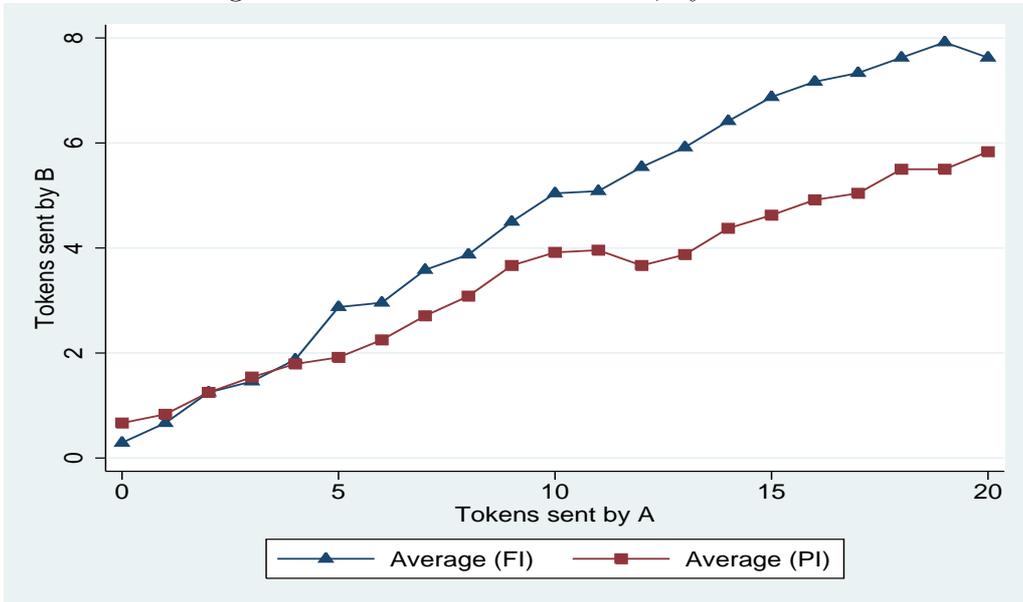


Figure 6: Median amount returned, by treatment

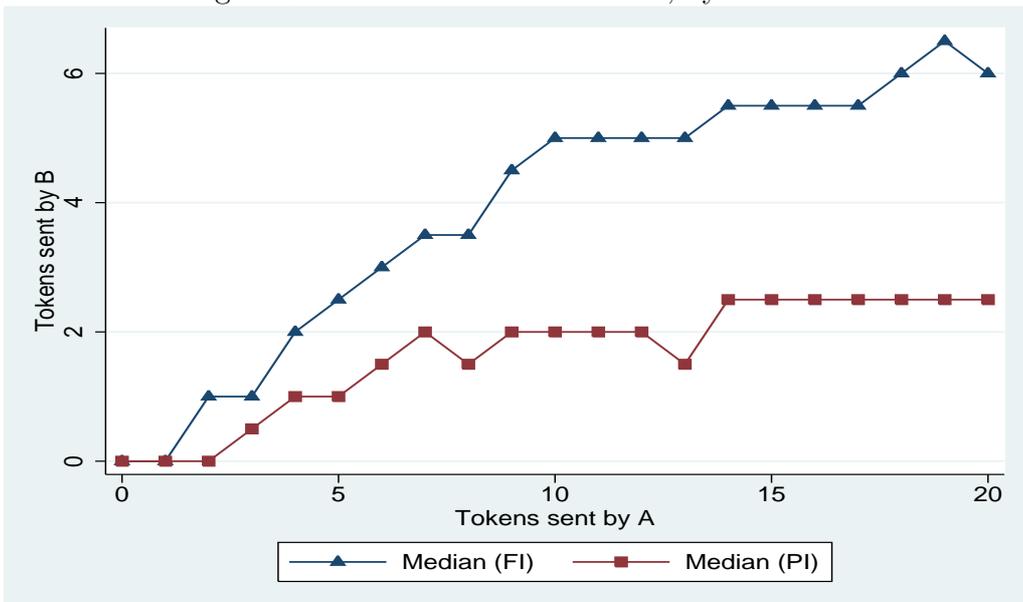


Figure 7: Individual responses of players B (SM): FI

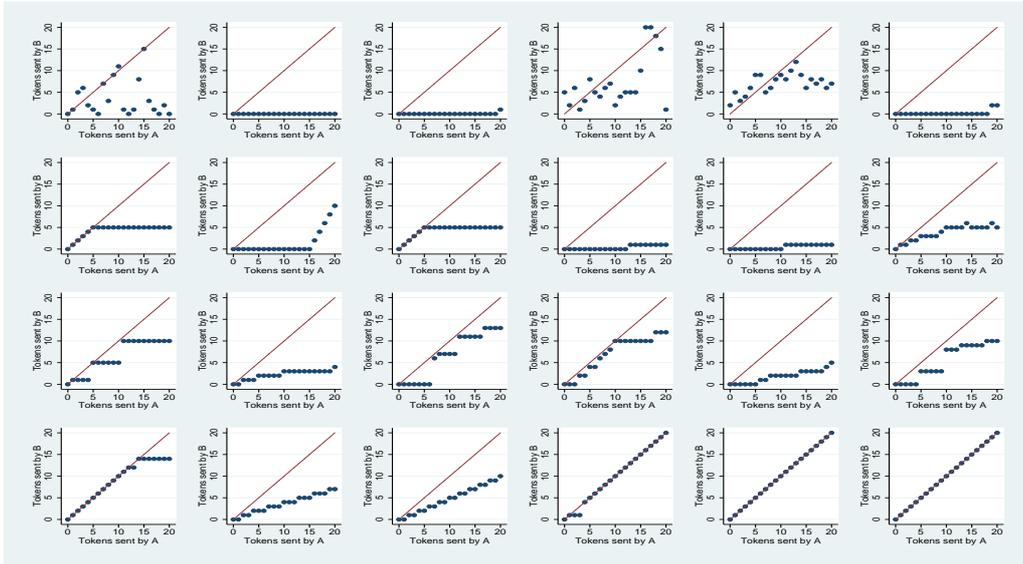


Figure 8: Individual responses of players B (SM): PI

