The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers
www.cesifo.org/wp

Does Instrumental Reciprocity Crowd out Prosocial Behavior?

Åshild A. Johnsen
Ola Kvaløy

CESIFO WORKING PAPER NO. 5078
CATEGORY 13: BEHAVIOURAL ECONOMICS
NOVEMBER 2014

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5078*

Does Instrumental Reciprocity Crowd out Prosocial Behavior?

Abstract

In repeated games, it is hard to distinguish true prosocial behavior from strategic instrumental behavior. In particular, a player does not know whether a reciprocal action is intrinsically or instrumentally motivated. In this paper, we experimentally investigate the relationship between intrinsic and instrumental reciprocity by running a two-period repeated trust game. In the 'strategic treatment' the subjects know that they will meet twice, while in the 'non-strategic treatment' they do not know and hence the second period comes as a surprise. We find that subjects anticipate instrumental reciprocity, and that intrinsic reciprocity is rewarded. In fact, the total level of cooperation, in which trust is reciprocated, is higher in the non-strategic treatment. This indicates that instrumental reciprocity crowds out intrinsic reciprocity: If one takes the repeated game incentives out of the repeated game, one sees more cooperation.

JEL-Code: C910.

Keywords: reputation, reciprocity, crowding out, experiment.

Åshild A. Johnsen
University of Stavanger
Stavanger / Norway
ashild.a.johnsen@uis.no

Ola Kvaløy
University of Stavanger
Stavanger / Norway
ola.kvaloy@uis.no

For valuable comments and suggestions, we would like to thank Jim Andreoni, Kjell Arne Brekke, Alexander Cappelen, Gary Charness, Kristoffer Eriksen, Nicholas Feltovich, Stein Holden, Terrance Odean, Trond Olsen, Mari Rege, Arno Riedl, Bettina Rockenbach, Joel Sobel, Sigrid Suetens, Bertil Tungodden, Joel Watson, Ro'i Zultan, and participants at the 2012 North-American ESA meeting, the 9th IMEBE meeting, the 6th M-BEES meeting, the 7th Nordic Conference in Behavioral and Experimental Economics, the 3rd Xiamen University International Workshop, and the annual meeting of Norwegian economists. Financial support from the Norwegian Research Council is greatly appreciated.

1 Introduction

The proverb "You always meet twice in life" is a rule of conduct for the dependable business person. The message is that honesty and trustworthiness always pay off in the end. This payoff is expectably highest if you surprisingly meet again. If you act trustworthy in your first meeting, and you surprisingly meet twice, your kindness may be perceived as more credible and your partner will more likely regard you as trustworthy. If, on the other hand, you know you will meet again, your kindness may be perceived as strategic.

Strategic kindness is termed "instrumental reciprocity" in the literature and contrasts with non-strategic "intrinsic reciprocity" (see Sobel, 2005). Instrumental reciprocity is part of a repeated game strategy where agents sacrifice short term gains in order to sustain reputation and increase long-term payoff. Intrinsic reciprocity, on the other hand, implies a willingness to sacrifice material payoff, either by rewarding a kind action or by punishing a mean action.

In this paper we investigate how instrumental reciprocity and intrinsic reciprocity interact. The standard idea is that instrumental reciprocity and reputational concerns amplify the positive effects of intrinsic reciprocity (i.e. that they are complements). If some people in the distribution are "good types" with intrinsic reciprocal preferences, then selfish types can imitate reciprocal types when playing a game repeatedly. Kreps et al. (1982) show that the mere belief in the existence of reciprocal types may sustain cooperative play for a number of periods in the finitely repeated prisoner's dilemma. This has been supported experimentally by Andreoni and Miller (1993). Also a number of trust games / gift exchange experiments with labor market contexts indicate that reputational concerns amplify the efficiency enhancing effects of reciprocity, see e.g. Falk and Gächter (2002) and Fehr, Brown and Zehnder (2009).

However, instrumental reciprocity may also crowd out or substitute for trust and intrinsic reciprocity. If the players know that their game is played repeatedly, they do not know whether reciprocal behavior is instrumentally or intrinsically

motivated. Now, if the players care about the motives behind the actions, i.e. if they value intrinsic reciprocity higher than instrumental reciprocity, then strategic repeated game incentives may potentially undermine trust and intrinsic reciprocity.[1]

We investigate the relationship between instrumental and intrinsic reciprocity by running a two period version of Berg et al's (1995) well-known trust game, called the investment game by the authors. In this game, the trustor, called the sender, sends money to a trustee, called the responder. The money is tripled. The responder then decides how much money to return to the sender. We run two treatments. In both treatments, senders and responders meet each other twice. In the strategic treatment the players know they will meet twice, while in the non-strategic treatment they do not know - the second round comes as a surprise. Hence, in the non-strategic treatment, we have a repeated game without repeated game incentives. This enables us to study how instrumental reciprocity interacts with intrinsic reciprocity, or more precisely how strategic repeated game incentives affect the repeated game.

In order to fix ideas, we analyze the two period investment game by using a standard reputation/reciprocity model that does not incorporate the idea that some agents may value intrinsic reciprocity higher than instrumental reciprocity. In the model there are two types of responders: good types who are always trustworthy, i.e. who always reciprocate, and bad types who are always self-interested. The responder knows his own type, but the sender only knows the chances that a responder is of either type. This model can account for the complementarity effect discussed above; more trust and reciprocity in the first period of the strategic treatment than in the first period of the non-strategic treatment. The model can also account for lower rates of reciprocity in the second period

[1]There is evidence that the motives and intentions behind actions matter. People are more prone to reciprocate good actions if they are confident about the motivation or intentions. See Gouldner (1960) for an early analysis, and e.g. McCabe et al. (2003) and Charness and Levine (2007) for more recent experimental evidence. However, the literature has not explicitly addressed potential crowding out effects from instrumental reciprocity. The crowding out literature has mainly focused on the negative effect of monetary incentives (Gneezy and Rustichini, 2000a,b; Bohnet et al. 2001; Fehr and Rockenbach, 2003) and monitoring (Falk and Kosfeld, 2006; Dickinson and Villeval, 2008).

of the strategic treatment, since a sender may trust a bad type in period 2 who instrumentally reciprocated in period 1. However, the total level of cooperation, in which a pair of subjects trusts and reciprocates trust in both periods, is predicted to be (weakly) higher in the strategic treatment.

The main results are as follows: First, as expected, repeated game incentives work in the first period. There is more trust and trustworthiness in the first period of the strategic treatment than in the first period of the non-strategic treatment. Second, senders who are reciprocated in the first period, show significantly more trust in the second period if they are in the non-strategic treatment. Hence, reciprocal behavior in the first period of the non-strategic treatment is perceived as more credible than reciprocal behavior in the strategic treatment. Third, there are considerably higher levels of trustworthiness in the second period of the non-strategic treatment than in the second period of the strategic treatment. While the senders' rate of return is on average -22% in the strategic treatment, it is 28% in the non-strategic treatment. Fourth, and most interestingly, the total level of cooperation, in which trust is reciprocated in both periods, is higher in the non-strategic treatment. This cannot be explained by the model**,** and indicates that strategic instrumental reciprocity crowds out prosocial behavior: If the repeated game incentives are taken out of the repeated game, more reciprocity is observed.

Our results can illuminate a more subtle research question that has been addressed recently about the relationship between instrumental and intrinsic reciprocity. The idea from Kreps et al. (1982), that more cooperation in finitely repeated games is due to reciprocity imitation is, as mentioned above, supported by a number of experiments, such as Andreoni and Miller (1993) and Falk and Gächter (2002). People act strategically/ instrumentally by imitating reciprocal preferences in order to sustain cooperation. However, the existing evidence for reciprocity imitation is not clean. More cooperation in finitely repeated games than in one shot games may not be due to reciprocity imitation, or uncertainty about types, but to an inappropriate use of the infinitely repeated game logic (see Rueben and Suetens, 2011). According to the theory, people play as if it were an infinitely repeated game, but change strategy in the final round when

they realize that the game will end. Hence, they act instrumentally, but they do not necessarily imitate reciprocal preferences or form beliefs about opponents' types. Experiments that find no differences between one shot games and the last round of the finitely repeated game are also supported by such a model of bounded rationality, and therefore cannot identify reciprocity imitation. Our experimental design makes it possible to identify reciprocity imitation. The reason is that players in the non-strategic treatment have received a more credible signal about the opponents' type prior to the last period. If they act on this signal in the last period, then reciprocity imitation is anticipated.

Related literature: A number of papers (including those cited above) experimentally investigate behavior in repeated games.[2] But in most of these experiments it is not possible to distinguish strategically from non-strategically motivated cooperation. Some recent papers aim to fill this gap. Ambrus and Pathak (2011) identify strategic cooperation in a finitely repeated public good game. By using a probabilistic continuation design, they show that selfish players do contribute to the public good because it induces future contributions by others. Reuben and Suetens (2012) investigate motives for cooperation in a repeated prisoner's dilemma game by using the strategy method.[3] Subjects *condition* their decisions on whether the period they are currently playing is the last period of the game or not. They find that most of the cooperation is strategically motivated. Cabral et al. (2012) identifies instrumental reciprocity by running an infinitely repeated veto game which admits a unique efficient equilibrium for selfish rational players. Like Rueben and Suetens they find that most of the cooperation is motivated by instrumental reciprocity.

The approach we use in this paper is related to a procedure called "surprise restart", an approach first reported by Andreoni (1988) which has been used in several papers since then.[4] Closest to our paper are Ben-Ner et al. (2004) and

[2]For an early experimental investigation of a finitely repeated trust game, see Camerer and Weigelt (1988). More recent papers are Anderhub et al. (2002), Engle-Warnick and Slonim (2004, 2006), and Schniter et al. (2013).

[3]In the strategy method for eliciting choices, subjects state their contingent choices for every decision node they may face. They are then matched, and their choices are played out.

[4]See Ambrus and Pathak (2011) and Rietz et al. (2013) for a recent application of the surprise

Stanca et al. (2009). Ben-Ner et al. run a two-part dictator game where the subjects are told *after* they have played a conventional dictator game that they will play one more time, with reversed roles. Those subjects who were treated nicely in the first part tended to be more generous towards the former dictators. Stanca et al. use a similar approach with the difference being that money sent from the dictator is tripled. Hence, when the roles are reversed, the game resembles the second stage of a standard gift exchange game. Treatments differ in whether or not the subjects are told about the second stage before the experiment starts. They find that reciprocity is stronger when strategic motivations can be ruled out. However, neither Ben-Ner et al. nor Stanca et al. can follow the behavior of a given subject (with a given role) over two periods, like we can. Hence, they do not study a repeated game, which is the object of our investigation.

The rest of the paper is organized as follows: In Section 2 we present a formal analysis of the two period investment game. In Section 3 we present the experimental design and procedure. In Section 4 we present the results, while Section 5 concludes.

2 The two period investment game

The investment game of Berg et. al (1995) is a trust game where the trustor (sender) chooses the degree to which she trusts the trustee (responder). The sender and responder start out with endowments of E_s and E_r, respectively. The sender chooses an amount x to send to the responder ($0 \leq x \leq E_s$). The investment x is then multiplied by $m > 1$ and the responder receives mx. Subsequently, the responder chooses an amount y that he or she returns to the sender, with $0 \leq y \leq mx$. Afterwards, the game ends with the sender receiving a payoff $E_s - x + y$ and the responder receiving $E_r + mx - y$.

restart method. Our approach is less of a surprise since we explicitly state that the experiment will consist of two parts, and that the instructions for the second part will come after the first part. On a more general level, our methodology is in line with experiments in which subjects are not informed about the content of all stages ex ante, since this might induce strategic behavior, see e.g. Dohmen and Falk (2011) and Bartling et al. (2012).

If $x > 0$ the sender (to some extent) trusts the responder. If $y > x$ the responder reciprocates, i.e. she is trustworthy. If both $x > 0$ and $y > x$ we say that the parties cooperate. The level of x indicates how much the sender trusts the responder, while the level of $y > x$ indicates how trustworthy the responder is.

Consider now the game played twice, with x_i, y_i, as choice variables and where $i = 1, 2$ denotes periods. With standard (selfish) preferences and common knowledge about rationality, the Nash equilibrium profile is ($x_1 = x_2 = 0$, $y_1 = y_2 = 0$). From a number of experiments on finitely repeated games, we know that this is not a plausible outcome.

We thus make the following assumptions: There are two types of responders. B-types (bad types) play $y_2 = 0$. G-types (good types) play $y_i = kx_i$ where $k > 1$. (Hence k is a definition of what a good type is, and thus exogenous). The responder knows his own type, but the sender only knows the chances that a responder is of either type. The distribution of types is common knowledge.

We will now sketch the derivation of a sequential equilibrium in the two period investment game.[5] In subgames along the equilibrium path, players are assumed to use Bayes' rule to update their information about others based on observed play. We solve by backwards induction:

In period 2, the sender knows that a b-type responder will play $y_2 = 0$ while a g-type responder will play $y_2 = kx_2$. Therefore, if the sender thinks the probability that the responder is a g-type is p_2,his expected payoff from sending a positive amount $x_2 > 0$ is $p_2(E_s - x_2 + kx_2) + (1 - p_2)(E_s - x_2)$. The payoff from sending exceeds the sure payoff from not sending (E_s) iff

$$p_2 > \frac{1}{k} \tag{1}$$

In period 1 a b-type responder can play $y_1 = 0$ and get $E_r + mx_1$ in period 1 and E_r in period 2. Alternatively, he can play a mixed strategy, choosing

[5] See Camerer and Weigelt (1988) for a similar analysis of t period trust game.

to reciprocate with probability σ_1 and not reciprocate with probability $1 - \sigma_1$. The responder will choose σ_1 so that when the sender observes the responder reciprocate in period 1, the sender updates his beliefs about the responder such that the updated posterior probability p_2 is above the threshold $\frac{1}{k}$. Then the responder's total expected payoff from periods 1 and 2 is:

$$\sigma_1(E_r + mx_1 - kx_1 + E_r + mx_2) + (1 - \sigma_1)(E_r + mx_1 + E_r) \quad (2)$$

Since this expected payoff is increasing in σ_1 (as long as $kx_1 < mx_2$), the responder will choose σ_1 as large as possible, provided σ_1 makes the posterior probability p_2 above the sender's threshold $\frac{1}{k}$. If the sender uses Bayes rule to update probabilities, the posterior probability p_2 is given by

$$p_2 = \frac{p_1}{[p_1 + \sigma_1(1 - p_1)]} \quad (3)$$

Note that for $\sigma_1 = 1$, i.e. a b-type always reciprocates in period 1, then $p_2 = p_1$, i.e. from observing reciprocal behavior, the sender does not increase his probability that he actually meets a good type. But for $\sigma_1 < 1$ reciprocity is a positive signal.

For p_2 to exceed the threshold $\frac{1}{k}$, we must have $\frac{p_1}{[p_1+\sigma_1(1-p_1)]} \geq \frac{1}{k}$ i.e.

$$\sigma_1 \leq \frac{p_1\,(k-1)}{(1-p_1)} \quad (4)$$

A rational responder will choose σ_1 so that (4) holds with equality.

Now, in period 1 the sender will send $x_1 > 0$ iff

$$[p_1 + \sigma_1(1 - p_1)]\,(E_s - x_1 + kx_1) + (E_s - x_1)(1 - \sigma_1)(1 - p_1) > E_s \quad (5)$$

Since the responder will choose σ_1 to satisfy (4), we can combine (4) and (5) to derive the threshold for p_1, which is

$$p_1 > \frac{1}{k^2} \tag{6}$$

We assume that the game begins with a commonly known prior probability h that the responder is a g-type. The sequential equilibrium is then for the sender to send and for the responder to reciprocate as long as the prior h is above the threshold. But if the responder sees in period 1 that the threshold for sending in period 2 ($h > \frac{1}{k}$) will be violated, the responder starts playing mixed strategies with probabilities to reciprocate as given by $\sigma_1 = \frac{p_1(k-1)}{(1-p_1)}$. Once mixed strategies begin, the responder's choice of σ_1 in equilibrium makes the sender indifferent between strategies and vice versa.[6]

Assume now that the game is played twice, but that they are informed about period 2 after period 1 is played. In this situation, instrumental reciprocity plays no role. In period 1 the sender knows that a b-type responder will play $y_1 = 0$ while a g-type responder will play $y_1 = kx_1$. Therefore, if the sender thinks the probability that the responder is a g-type is h, the payoff from sending exceeds the sure payoff from not sending iff $h > \frac{1}{k}$**.** The g-types reciprocate, while b-types do not. In period 2 the sender knows the responder's type (since there are no reputation building / no mixed strategies). He sends if he met a g-type in period 1 and does not send if he met a b-type. Hence, there is a probability h that he sends in period 2. Again, a g-type reciprocates, while a b-type does not. If $h < \frac{1}{k}$, then there is no trust (and hence no trustworthiness) in any of the two periods.

We can now compare the two situations. For convenience, let us call the standard two period game "the strategic game", and the game where period 2 comes

[6]The sender chooses to send with probability q, where q makes the responder indifferent between values of σ_1 in the following expression for his expected payoff: $\sigma_1(E_r + mx_1 - kx_1 + (E_r + mx_2)q + E_r(1-q)) + (1-\sigma_1)(E_r + mx_1 + E_r)$.When $q = \frac{k}{m}$, this expression is equal to the responder's expected payoff when the sender does not mix (2).

as a surprise "the non-strategic game". First, we see that the model can account for complementarity between instrumental and intrinsic reciprocity: The possibility of meeting good types makes it possible for bad types to imitate good types in period 1 in the strategic game. Hence, the level of trust and trustworthiness is weakly higher in period 1 of the strategic game than in period 1 of the non-strategic game. In period 2, the trust level is (weakly) higher in the strategic game, since if $h > \frac{1}{k}$, the sender sends in both periods in the strategic game, while he only sends with probability h in period 2 of the non-strategic game. With respect to trustworthiness, only a sender that met a good type in period 1 of the non-strategic game will subsequently trust in period 2. By contrast, in the strategic game, a sender may meet a bad type in period 2 who instrumentally reciprocated in period 1. Hence, the likelihood of being reciprocated, conditional on sending a positive amount, is higher in period 2 of the non-strategic game. However, since the probability of initially meeting a good type is equal in the two games, and the trust level is higher in the strategic game, the total level of cooperation - in which a pair of subjects trusts and reciprocates trust (respectively) in both periods - is (weakly) higher in the strategic game.

3 Experimental design and procedure

We run an experiment in which subjects play the two period investment game analyzed in the previous section. In one treatment, which we denote the strategic treatment, subjects knew they were going to meet twice. In the other treatment, denoted non-strategic treatment, the second period came as a surprise.

The subjects play four one shot versions of the game each period. More specifically, we announced that the experiment consisted of *two parts*. In the first part they were going to play the investment game four times, which we will refer to as *rounds*, each round against a new opponent. In the second and last part, subjects were going to play four rounds against the same four opponents they met in the first part. Prior to each round of the second part, they got information about how they played when they met each other in Part I.

The two treatments were identical except for when information was revealed. In the strategic treatment all information was revealed prior to the first part. In the non-strategic treatment we announced that the experiment consisted of two parts, then explained only the first part and said that information about the second and last part would be given after the first part was finished.

In the beginning of the experiment subjects were assigned the role as a sender or a responder, and they kept the same role throughout the experiment. Senders and responders were randomly paired in each round of the first part. In the t^{th} round of the second part the sender met the same responder as in the t^{th} round of the first part. In each round subjects were endowed with 100 ECU each $(100ECU = 20NOK \approx \$3.5)$. The sender was then given the opportunity to send an amount x from her endowment to the responder. The amount of money sent by the sender was tripled $(m = 3)$ by the experimenter so that the responder received $3x$. Then the responder had the opportunity to send back to the sender an amount y. Hence, in a given round, the sender's payoff was $100 - x + y$, while the responder's payoff was $100 + 3x - y$.

The experiment was conducted in March 2012 at the University of Stavanger, Norway. In all 196 subjects participated, 102 in the strategic treatment and 94 subjects in the non-strategic treatment. Average earning per subject was $\$43$. All instructions were given both written and verbally. The experiment was conducted and programmed with the software z-Tree (Fischbacher 2007).

4 Results

We start by investigating the average trust and trustworthiness levels, and the average sent and returned amounts. We then look into how the subjects' behavior is conditional on their opponents' actions, and finally we study how cooperation evolves.

Table 1 presents the average trust and trustworthiness rates for part I and part II. The trust rate is defined as the fraction of individual senders who send a positive amount. The trustworthiness rate is defined as the fraction of individual responders who return more than what they received - conditional on the senders sending them anything.

		Trust rate		Trustworthiness rate	
	Ind	Part I	Part II	Part I	Part II
Strategic	51	0.98	0.77	0.73	0.34
Non-strategic	47	0.88	0.70	0.53	0.62

Table 1: Trust and trustworthiness rates (individual averages)

First we see that both the trust rate and the trustworthiness rate are highest in part I of the strategic treatment. Trust and trustworthiness are significantly higher in the first part of the strategic treatment compared to the first part of the non-strategic treatment. We also see that both the trust and the trustworthiness rates are significantly reduced from part I to II in the strategic treatment.[7] This is all predicted by the model.

Result 1: *The rates of trust and trustworthiness are highest in part I of the strategic treatment.*

Next consider part II. We see a slightly higher trust rate in the strategic treatment, but it is not significant (p=*0.23*). With respect to trustworthiness, we see, as expected, a higher rate of trustworthiness in the non-strategic treatment. If there are subjects who act instrumentally and imitate reciprocal preferences, then the likelihood of not being reciprocated is lower in part II of the strategic game.

[7] We calculate the average trust and trustworthiness rates of each individual in the two treatments. Mann-Whitney tests on these averages reveal the following p-values (two-tailed) between treatments: trust part I p=0.01, trust part II p=0.23. Between treatments: Trustworthiness part I p=0.01, trustworthiness part II $p<0.01$. Furthermore, we compare part I to part II in each treatment (Wilcoxon signed-rank test): Trust in the strategic treatment $p<0.01$, trust in the non-strategic treatment: $p<0.01$. Trustworthiness in the strategic treatment $p<0.01$, trustworthiness in the non-strategic treatment p=0.34.

Result 2: *There are higher rates of trustworthiness in part II of the non-strategic treatment than in part II of the strategic treatment.*

We now turn to the *levels* of trust and trustworthiness. Before considering the regressions, it is useful to look at the summary statistics. Table 2 presents the average amount sent and returned, in addition to rate of return (RoR)[8], for those senders (responders) who chose to trust (were trusted).[9]

	Part I			Part II		
	Sent	Ret	RoR	Sent	Ret	RoR
Strategic	60.9	96.1	0.48	57.9	51.2	-0.22
Non-strategic	71.2	80.0	0.05	73.8	90.3	0.28

Table 2: Average sent, returned amounts, and rate of return for those who exhibited/ experienced trust (individual averages).

The senders who chose to trust in the strategic treatment send on average about 60 percent of their endowment, while the senders in the non-strategic treatment send about 70 percent in both part I and part II. This difference is only statistically different in part II. We can see from Table 2 that the responders in the strategic treatment return more in part I, and significantly less in part II compared to the non-strategic treatment. The responders in the strategic treatment reduce how much they return significantly, while there is no difference between amount returned in the first and the second part for the non-strategic treatment. This is reflected in the senders' rates of return. In the strategic treatment, the senders initially experience a significantly higher rate of return (0.48), which is reduced dramatically and significantly in part II, and actually turns out negative (-0.22). In the non-strategic treatment we observe the opposite pattern: from an initial low but positive rate of return, the senders' payoff from sending increases significantly in the final part.[10]

[8] Rate of return is a simple measure indicating the payoff from trusting for the senders, and it is the difference between sent and returned amount over sent amount.

[9] Summary statistics for all decisions can be found in the appendix.

[10] We calculate the average sent amount, returned amount and rate of return for each individual in the two treatments, for those subjects who exhibited and experienced trust. Mann-Whitney

We now look closer at how trust, trustworthiness and cooperation in part II depend on previous behavior. First, we investigate how the senders' trust levels in part II are conditional on the level of trustworthiness experienced in part I. Table 3 presents the results from four regressions, where sent amount in part II is regressed on a dummy variable for being in the non-strategic treatment, amount returned in part I, and an interaction between the non-strategic treatment and amount returned in part I. In addition we control for the senders' age, gender, faculty background, and possible round effects (we report robust clustered standard errors for individuals).

---

tests on these averages reveal the following p-values (two-tailed) between treatments: Sent part I p=0.17, sent part II p=0.01. Between treatments: returned part I p=0.16, returned part II $p<0.01$. Between treatments: RoR part I $p<0.01$, RoR part II $p<0.01$. We also compare part I and part II using Wilcoxon signed-rank tests. Part I and part II Strategic treatment: Sent p=0.31, returned $p<0.01$, RoR $p<0.01$. Part I and part II non-strategic treatment: Sent p=0.14, returned p=0.21, RoR $p<0.01$.

Sent in part II	(a)	(b)	(c)	(d)
Non-strategic treatment	-0.32	0.06	12.36***	2.53
	(6.218)	(4.531)	(4.222)	(5.944)
Sent part I		0.50***	-0.08	-0.08
		(0.065)	(0.092)	(0.091)
Returned amount part I			0.42***	0.37***
			(0.047)	(0.057)
Returned part I*treatment				0.12**
				(0.056)
Dummy: male	19.07**	9.67*	6.14	5.70
	(7.369)	(5.794)	(5.305)	(5.294)
Age	1.04	0.95	0.67	0.60
	(0.675)	(0.573)	(0.476)	(0.504)
Science and technology	1.62	-1.05	2.21	1.71
	(7.189)	(5.318)	(4.756)	(4.782)
Arts and education	-7.69	-3.82	-2.12	-2.16
	(8.551)	(6.968)	(6.325)	(6.536)
Round 2	-4.85	-8.01*	-9.44**	-9.03**
	(4.483)	(4.770)	(4.012)	(3.936)
Round 3	-13.83***	-20.34***	-17.54***	-16.79***
	(4.848)	(5.219)	(4.003)	(3.897)
Round 4	-12.59***	-17.79***	-14.87***	-14.11***
	(4.214)	(4.655)	(3.941)	(3.865)
Constant	24.30	3.10	4.49	10.46
	(14.953)	(13.750)	(11.095)	(11.897)
Adjusted R^2	0.079	0.254	0.460	0.469
Observations	392	392	392	392

Table 3: Sent in part II (OLS), robust clustered standard errors (individuals), * 0.10 ** 0.05 *** 0.01.

We see that regressions (a) and (b) imply no net treatment effect. When we control for how much the responders return in part I in regression (c), we see that the senders in the non-strategic treatment seem to trust more than the senders in the strategic treatment in part II. In order to investigate whether experiencing trustworthy behavior in part I has a stronger effect on trust in part II for the subjects in the non-strategic treatment, we include an interaction term between returned amount in part I and being in the non-strategic treatment in (d). We see that trust that is rewarded, i.e. reciprocated, has a stronger effect in the non-strategic treatment. We have:

Result 3: *Senders who are reciprocated in part I show more trust in part II if*

they are in the non-strategic treatment.

Result 3 implies that reciprocal behavior in the first part of the non-strategic treatment is perceived as more credible than reciprocal behavior in the strategic treatment. This shows that subjects anticipate instrumental reciprocity in the strategic treatment.

Let us now look closer at the responders. We have seen that on average there are higher rates of trustworthiness in part II of the non-strategic treatment than in part II of the strategic treatment. From Table 4 we also see that responders return a larger share in part II when we control for senders' behavior and background variables.

Percentage returned part II	(1)	(2)
Non-strategic treatment	13.68***	11.99***
	(4.501)	(4.452)
Sent part I		0.09
		(0.053)
Sent part II		0.09
		(0.067)
Dummy: male	-3.14	-3.27
	(5.287)	(5.139)
Age	0.02	-0.03
	(0.424)	(0.404)
Science and technology	-7.36	-7.89
	(5.135)	(5.095)
Arts and education	-5.37	-4.13
	(6.960)	(6.746)
Round 2	-3.43	-4.12
	(2.834)	(2.811)
Round 3	-1.25	-2.11
	(3.083)	(3.445)
Round 4	1.29	0.23
	(3.098)	(3.230)
Constant	33.39***	24.44**
	(10.840)	(10.364)
Adjusted R^2	0.091	0.131
Observations	289	289

Table 4: Percentage returned by responders in part II (OLS), robust clustered standard errors (individuals), * 0.10 ** 0.05 *** 0.01.

The linear probability models in Table 5 further illuminate the cooperative behavior in part II.[11] Cooperation means that the sender chooses to trust, and the responder reciprocates. The dependent variable is whether the subjects chose to cooperate or not in part II - cooperation is a dummy variable equal to one for each pair where trust was reciprocated. In (I) we regress the probability of cooperation in part II on a dummy for being in the non-strategic treatment, with the same controls as before.

Cooperation part II	(I)	(II)	(III)	(IV)
Non-strategic treatment	0.16***	0.27***	0.28***	0.09
	(0.049)	(0.051)	(0.049)	(0.064)
Constant	0.16	0.05	-0.12	-0.01
	(0.143)	(0.142)	(0.144)	(0.148)
Sent part I		-0.00**	0.00	0.00
		(0.001)	(0.001)	(0.001)
Returned amount part I		0.00***	0.00	0.00
		(0.001)	(0.001)	(0.001)
Cooperation part I			0.34***	0.19**
			(0.072)	(0.078)
Cooperation part I*treatment				0.29***
				(0.086)
Dummy: male	0.04	-0.02	-0.03	-0.04
	(0.053)	(0.056)	(0.054)	(0.053)
Age	0.01	0.01	0.01	0.01
	(0.006)	(0.006)	(0.005)	(0.005)
Science and technology	-0.04	-0.04	-0.02	-0.02
	(0.054)	(0.053)	(0.051)	(0.051)
Arts and education	-0.13	-0.11	-0.10	-0.11
	(0.078)	(0.080)	(0.078)	(0.075)
Round 2	-0.08	-0.10	-0.10	-0.10
	(0.069)	(0.069)	(0.066)	(0.064)
Round 3	-0.07	-0.07	-0.08	-0.07
	(0.068)	(0.068)	(0.066)	(0.065)
Round 4	-0.05	-0.05	-0.07	-0.06
	(0.069)	(0.068)	(0.066)	(0.066)
Adjusted R^2	0.029	0.144	0.196	0.214
Observations	392	364	364	364

Table 5: Cooperation in part II (linear probability model), robust standard errors, * 0.10 ** 0.05 *** 0.01.

[11] For samples of this size, linear probability models are more robust than logit or probit models.

The treatment dummy shows that it is more likely that a pair cooperates in the second part of the non-strategic treatment. Controlling for how much the sender sent in part I and how much the responder returned in part I, we see from (II) that the point estimate suggests an even higher probability of cooperation in the non-strategic treatment. In (III) we add a dummy for whether or not they cooperated the first time they met, and this dummy captures the effects from sent and returned amounts. The treatment effect remains strong and significant. In (IV) we add an interaction term between cooperation in part I and being in the non-strategic treatment, and we see that the probability of cooperation - in which trust is reciprocated - is higher in the non-strategic treatment for those subjects who have experienced cooperation in part II. We have:

Result 4: *The probability of cooperation in part II of the non-strategic treatment is higher than in part II of the strategic treatment.*

Finally, we are interested in the total level of cooperation during the repeated game. Figure 1 compares the rate of pairs cooperating between parts and treatments.

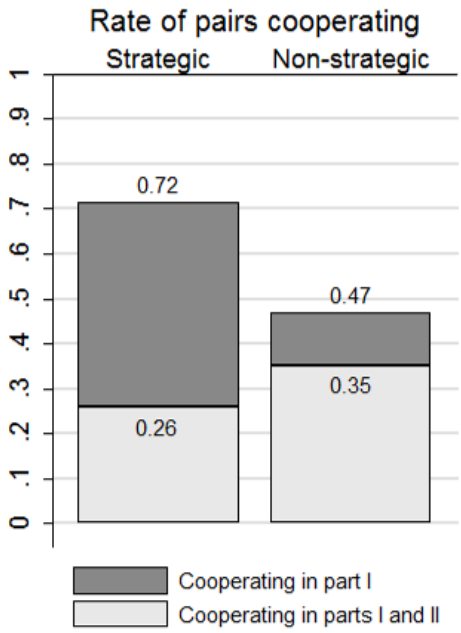

Figure 1: Rate of pairs cooperating.

In the strategic treatment, 72 percent (146 of 204 observations) of the pairs cooperate in the first part. This is significantly higher than the first part of the non-strategic treatment. 26 percent of the pairs cooperate in both part I and part II of the experiment, which is significantly less than in part I. In the non-strategic treatment 47 percent (88 of 188) cooperate in part one, and 35 percent cooperate in both part I and part II. So while the initial cooperation rates are significantly larger in the first part of the strategic treatment (72 versus 47), the rate of pairs who cooperate *throughout both parts* is significantly larger than that of the strategic treatment (26 versus 35).[12] We have:

Result 5: *The total level of cooperation, in which trust is reciprocated in both parts, is higher in the non-strategic treatment.*

[12] We compare the cooperation rates for each pair. Mann-Whitney tests (two-tailed) between treatments: p for part 1, p=0.05 for part II. Wilcoxon signed-rank test comparing parts I and II for the non-strategic and strategic treatments, respectively: $p<0.01$, $p<0.01$.

This result cannot be explained by the standard reputation/reciprocity model outlined in Section 2, and shows that instrumental reciprocity can crowd out true prosocial behavior.

5 Concluding remarks

In repeated games it is hard to distinguish intrinsic reciprocity from strategic instrumental reciprocity. As a result, the latter can substitute for the former. We investigate this by running a two period version of Berg et al's (1995) well-known trust game, also called the investment game. We run two treatments. In the strategic treatment the players know they will meet twice, while in the non-strategic treatment they do not know - the second period comes as a surprise. Hence, in the non-strategic treatment we have a repeated game without repeated game incentives. This enables us to study how instrumental reciprocity interacts with intrinsic reciprocity, or more precisely how strategic repeated game incentives affect the repeated game.

We find that subjects who are reciprocated in the first period (part), show significantly more trust in the second period if the second period comes as a surprise. This implies that subjects anticipate instrumental reciprocity in the strategic treatment. Moreover, we find that the total level of cooperation, in which trust is reciprocated in both periods, is higher in the non-strategic treatment. Hence, if one takes the repeated game incentives out of the repeated game, one sees more cooperation.

Our paper thus provides evidence that instrumental reciprocity may, under given conditions, crowd out intrinsic reciprocity. However, we cannot identify the exact mechanism behind the crowding out result. The problem with the strategic treatment is that subjects can neither reward intrinsic reciprocity (relevant for the senders), nor signal intrinsic reciprocity (relevant for the responders). Future research should try to disentangle these two potential sources for the crowding out result.

References

[1] Ambrus, Attila, and Parag A. Pathak, 2011. Cooperation over Finite Horizons: A Theory and Experiments. *Journal of Public Economics*, 95(7-8): 500-512.

[2] Anderhub,Vital, Engelmann, Dirk, and Werner Güth. 2002. An Experimental Study of the Repeated Trust Game with Incomplete Information, *Journal of Economic Behavior & Organization*, 48(2): 197–216.

[3] Andreoni, James. 1988. Why free ride? Strategies and Learning in Public Goods Experiments, *Journal of Public Economics*, 37(3): 291–304.

[4] Andreoni, James, and John H. Miller. 1993. Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence. *Economic Journal*, 103: 570–585.

[5] Bartling, Björn, Fehr, Ernst, and Klaus M. Schmidt. 2012. Screening, Competition, and Job Design: Economic Origins of Good Jobs. *American Economic Review*, 102(2): 834–64.

[6] Ben-Ner, Avner, Putterman, Louis, Kong, Fanmin and Dan Magan. 2004. Reciprocity in a Two-part Dictator Game. *Journal of Economic Behavior & Organization*, 53(3): 333–352.

[7] Berg, Joyce , Dickhaut, John, and Kevin McCabe. 1995. Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10(1): 122–142.

[8] Bohnet, Iris, Frey, Bruno S., and Steffen Huck. 2001. More Order with Less Law: On Contract Enforcement, Trust and Crowding. *American Political Science Review*, 95(1): 131–144.

[9] Cabral, Luis M. B., Erkut Ozbay, and Andrew Schotter. 2012. Intrinsic and Instrumental Reciprocity: An Experimental Study. *Working Paper*.

[10] Camerer, Colin, and Keith Weigelt. 1988. Experimental Tests of a Sequential Equilibrium Reputation Model. *Econometrica*, 56(1): 1–36.

[11] Charness, Gary, and David I. Levine. 2007. Intention and Stochastic Outcomes: An Experimental Study, *Economic Journal*, 117(522): 1051–1072.

[12] Dickinson, David, and Marie-Claire Villeval. 2008. Does Monitoring Decrease Work Effort? The Complementarity between Agency and Crowding-Out Theories, *Games and Economic Behavior*, 63(1): 56-76.

[13] Dohmen, Thomas, and Armin Falk. 2011. Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *American Economic Review*, 101(2): 556–590.

[14] Engle-Warnick, Jim, and Robert L. Slonim. 2004. The Evolution of Strategies in a Repeated Trust Game. *Journal of Economic Behavior & Organization*, 55(4): 553–573.

[15] Engle-Warnick, Jim, and Robert L. Slonim. 2006. Inferring Repeated-game Strategies from Actions: Evidence from Trust Game Experiments. *Economic Theory*, 28(3): 603–632.

[16] Falk, Armin, Fehr, Ernst, and Urs Fischbacher. 2008. Testing Theories of Fairness - Intentions Matter. *Games and Economic Behavior*, 62: 287–303.

[17] Falk, Armin, and Michael Kosfeld. 2006. The Hidden Costs of Control. *American Economic Review*, 96(5): 1611–1630.

[18] Falk, Armin, and Simon Gächter. 2002. Reputation and Reciprocity: Consequences for the Labour Relation. *Scandinavian Journal of Economics*, 104: 1–26.

[19] Fehr, Ernst, Brown, Martin, and Christian Zehnder. 2009. On Reputation: A Microfoundation of Contract Enforcement and Price Rigidity. *Economic Journal*, 119(536): 333–353.

[20] Fehr, Ernst, and Urs Fischbacher. 2002. Why Social Preferences Matter - the Impact of Non-Selfish Motives on Competition, Cooperation and Incentives. *Economic Journal*, 112(478): C1-C33.

[21] Fehr, Ernst, and Bettina Rockenbach. 2003. Detrimental Effects of Sanctions on Human Altruism. *Nature*, 422: 137–140.

[22] Fischbacher, Urs. 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10: 171–178.

[23] Gneezy, Uri, and Aldo Rustichini. 2000. Pay Enough or Don't Pay At All. *Quarterly Journal of Economics*, 115(3): 791–810.

[24] Gneezy, Uri, and Aldo Rustichini. 2000. A Fine is a Price, *Journal of Legal Studies*, 29(1): 1–17.

[25] Gouldner, Alvin W. 1960. The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review*, 25(2): 161–178.

[26] Kreps, David M., Milgrom, Paul, Roberts, John, and Robert Wilson. 1982. Rational Cooperation in the Finitely Repeated Prisoners' Dilemma, *Journal of Economic Theory*, 27(2): 245–252.

[27] McCabe, Kevin A., Rigdon, Mary L., and Vernon L. Smith. 2003. Positive Reciprocity and Intentions in Trust Games, *Journal of Economic Behavior & Organization,* 52(2): 267–275.

[28] Reuben, Ernesto, and Sigrid Suetens. 2011. Maladaptive Reciprocal Altruism. *Working Paper Columbia University*.

[29] Reuben, Ernesto, and Sigrid Suetens, 2012. Revisiting Strategic versus Non-Strategic Cooperation, *Experimental Economics*, 15: 24–43.

[30] Rietz, Thomas A., Sheremeta, Roman M., Shields, Timothy W., and Vernon L. Smith, 2013. Transparency, Efficiency and the Distribution of Economic Welfare in Pass-Through Investment Trust Games, *Journal of Economic Behavior & Organization*, 94: 257–267.

[31] Schniter, Eric, Sheremeta, Roman M., Sznycer, Daniel. 2013. Building and Rebuilding Trust with Promises and Apologies, *Journal of Economic Behavior & Organization*, 94: 242–256.

[32] Sobel, Joel. 2005. Interdependent Preferences and Reciprocity. *Journal of Economic Literature*, 43(2): 392–436.

[33] Stanca, Luca, Bruni, Luigino, and Luca Corazzini. 2009. Testing Theories of Reciprocity: Do Motivations Matter? *Journal of Economic Behavior & Organization*, 71(2): 233–245.

6 Appendice

6.1 Summary statistics

	Part I				Part II			
	Decisions	Sent	Ret	RoR	Decisions	Sent	Ret	RoR
Strategic	204	60.7	93.5	0.49	204	46.3	41.9	-0.16
Non-strategic	188	63.8	70.1	0.07	188	50.7	66.1	0.19

Table 6: Average sent, returned amounts, and rate of return, all observations.

6.2 Rounds

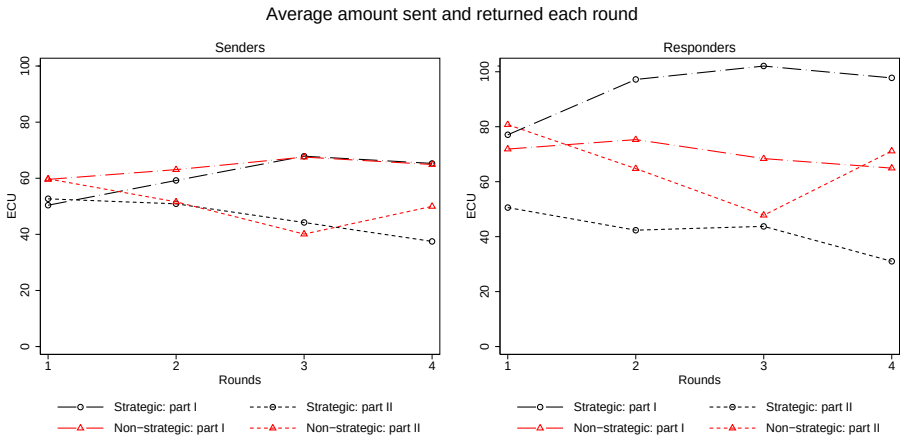

Figure 2: Average amount sent and returned in each round.

Instructions for the strategic treatment:

Welcome!

This experiment will last for about 30 minutes. Throughout the experiment you will get the opportunity to earn money that will be paid out in cash and anonymously after the experiment is over.

You will now be given time to read through the instructions for this experiment. If you have any questions concerning these instructions, please raise your hand and we will come to you. Talking or communicating with others is not allowed during the experiment.

Throughout this experiment we will use experimental kroner (EK), and not Norwegian kroner (NOK). By the end of the experiment, the total amount of EK which you have earned will be converted to NOK at the following rate*:*

5 EK = 1 NOK

6.2.1 Instructions

This experiment consists of two parts, Part I and Part II – and these parts are identical.

All subjects are divided into pairs, where one will be a sender and one will be a responder. This means that half of you will get to be senders and half of you responders. You will not get to know who will be your partner. Your partner is

in the room, but you will not get to know who this person is, neither during nor after the experiment.

Each part consists of four rounds. In each round of Part I a sender meets a new responder.

In Part II senders and responders meet again in the same order that they met in Part I. In other words: the person you met in round 1 of Part I, you will meet again in round 1 of Part II. The person you met in round 2 of Part I, you will meet again in round 2 of Part II, and so on.

PART I:

ROUND 1: In the beginning of each round all participants receive 100 EC.

The sender can now choose to send everything, nothing or some of the 100 EC to the responder. The money which is sent is then tripled. If the sender chooses to send for instance 20 EC to the responder, the responder receives 60EC. If 90 EC is sent, the responder receives EC 270.

The responder then decides how much he/ she wants to keep, and how much he/ she would like to return. The money which will be sent back will not be tripled.

When the sender chooses to send an amount x of the100 EC to the responder and the responder returns y, the total income of the sender in each round then equals:

100-x+y

The responder receives three times the sent amount x, and returns y. In addition, he/ she has the 100 EC he received in the start. The total income for the responder in each round then equals:

100+3x-y

By the end of each round the income from that round is put into the participant's account, and the round ends.

ROUNDS 2, 3 AND 4: Rounds 2, 3 and 4 are identical to round 1. Remember that each sender meets a new responder in each round. By the end of each round the money earned will be put into each participant's account.

PART II: Rounds 1, 2, 3 and 4 are identical to rounds 1, 2, 3 and 4 in Part I. It means that the sender meets the same responder in each round who he met in Part I. Before each round in Part II you will be reminded of the choices you made when you met in Part I.

Please follow the messages which appear on the screen. In the end you will be asked to fill out a short questionnaire, and you will be informed about your total earnings converted into NOK.

On the pc cabinet you can see a white sticker with the logo of the university, and a number, for instance D10136. Please write down this number and your total income on the receipt when the experiment is over. When we tell you that the

experiment is over, you can leave the room with the receipt. Bring this to the EAL building, office H-161, to collect your total earnings.

Instructions for the non-strategic treatment:

Welcome!

This experiment will last for about 30 minutes. Throughout the experiment you will get the opportunity to earn money that will be paid out in cash and anonymously after the experiment is over.

You will now be given time to read through the instructions for this experiment. If you have any questions concerning these instructions, please raise your hand and we will come to you. Talking or communicating with others is not allowed during the experiment.

Throughout this experiment we will use experimental kroner (EK), and not Norwegian kroner (NOK). By the end of the experiment, the total amount of EK which you have earned will be converted to NOK at the following rate*:*

5 EK = 1 NOK

Instructions

This experiment consists of two parts, and you will now receive the instructions for Part I of the experiment. Part II we will explain to you later.

All subjects are divided into pairs, where one will be a sender and one will be a responder. This means that half of you will get to be senders and half of you responders. You will not get to know who will be your partner. Your partner is

in the room, but you will not get to know who this person is, neither during nor after the experiment.

Each part consists of four rounds. In each round of Part I a sender meets a new responder.

PART I:

ROUND 1: In the beginning of each round all participants receive 100 EC.

The sender can now choose to send everything, nothing or some of the 100 EC to the responder. The money which is sent is then tripled. If the sender chooses to send for instance 20 EC to the responder, the responder receives 60EC. If 90 EC is sent, the responder receives EC 270.The responder then decides how much he/ she wants to keep, and how much he/ she would like to return. The money which will be sent back will not be tripled. When the sender chooses to send an amount x of the100 EC to the responder and the responder returns y, the total income of the sender in each round the equals:

$$100-x+y$$

The responder receives three times the sent amount x, and returns y. In addition, he/ she has the 100 EC he received in the start. The total income for the responder in each round then equals:

$$100+3x-y$$

By the end of each round the income from that round is put into the participant's account and the round ends.

ROUNDS 2, 3 AND 4: Rounds 2, 3 and 4 are identical to round 1. Remember that each sender meets a new responder in each round. By the end of each round the money earned will put into each participant's account.

Please follow the messages which appear on the screen. In the end you will be asked to fill out a short questionnaire, and you will be informed about your total earnings converted into NOK. On the pc cabinet you can see a white sticker with the logo of the university, and a number, for instance D10136. Please write down this number and your total income on the receipt when the experiment is over. When we tell you that the experiment is over, you can leave the room with the receipt. Bring this to the EAL building, office H-161, to collect your total earnings.