

6-2005

When Punishment Fails: Research on Sanctions, Intentions and Non-Cooperation

Daniel Houser
George Mason University

Erte Xiao
University of Pennsylvania, exiao@andrew.cmu.edu

Kevin McCabe
George Mason University

Vernon Smith
George Mason University

Follow this and additional works at: <http://repository.cmu.edu/sds>

Published In

Games and Economic Behavior , 62, 2, 509-532.

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Social and Decision Sciences by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

**WHEN PUNISHMENT FAILS:
RESEARCH ON SANCTIONS, INTENTIONS AND NON-COOPERATION**

Forthcoming Games and Economic Behavior

DANIEL HOUSER^{a*} ERTE XIAO^{b*} KEVIN MCCABE^c VERNON SMITH^c

- a. Interdisciplinary Center for Economic Science, Department of Economics, and Center for the Study of Neuroeconomics, George Mason University, 4400 University Drive, MSN 1B2, Fairfax, VA. 22030. USA.
- b. Philosophy, Politics and Economics Program and Wharton School, 313 Logan Hall, 249 S 36th Street, University of Pennsylvania. Philadelphia, PA, 19104. USA

Abstract: People can become less cooperative when threatened with sanctions, and previous research suggests both "intentions" and incentives underlie this effect. We report data from an experiment aimed at determining the relative importance of intentions and incentives in producing non-cooperative behavior. Participants play a one-shot investment experiment in pairs. Investors send an amount to trustees, request a return on this investment and, in some treatments, can threaten sanctions to enforce their requests. Decisions by trustees facing threats imposed (or not) by investors are compared to decisions by trustees facing threats imposed (or not) by nature. When not threatened, trustees typically decide to return a positive amount less than the investor requested. When threatened this decision becomes least common. If the request is large relative to the sanction then most trustees return nothing. If the request is small, trustees typically return the requested amount. These results do not vary with investors' intentions.

JEL classification: C71; C91

Keywords: Punishment, Cooperation, Trust, Incentives, Intentions, Experiment

* Corresponding authors.

E-mail addresses: dhouser@gmu.edu (D. Houser), exiao@sas.upenn.edu (E. Xiao).

Acknowledgements: Houser and Xiao contributed equally to this project. We are grateful to Ernst Fehr and Bettina Rockenbach for providing data and instructions related to their 2003 paper, and for many useful comments on early drafts of this paper. We also received valuable comments from an associate editor, three anonymous referees, James Andreoni, Colin Camerer, Tyler Cowen, John Dickhaut, David Dickinson, Martin Dufwenberg, Read Montague, Francesco Parisi, Aldo Rustichini and Bart Wilson, as well as participants at the 2004 SEA meetings, 2005 ESA and SABE meetings, and seminars at Max Plank (Bonn), Vassar College, University of Mannheim, University of Munich, University of Toulouse, University of Pennsylvania, University of Texas at Dallas, Laval University, James Madison University, NIH/NIDA, the Interdisciplinary Center for Economic Science and George Mason University's graduate student workshops. We thank many ICES graduate students, and particularly Bridget Butkevich, for assistance with experiments. We gratefully acknowledge the International Foundation for Research in Experimental Economics and the National Science Foundation (SES-0129744 and SES-0339181) for funding that supported this research.

1. Introduction

Sanctions are widely used to promote compliance in exchange environments. For example, an employer might wish to implement a policy requiring a specific level of employee effort, with fines levied upon those who shirk. Of course, threats are not necessarily effective. Under threats employees might shirk less yet still more than directed, or perhaps shirk even more than they had previously. Avoiding such inefficient outcomes is a matter of substantial practical importance, and requires developing a thorough understanding regarding conditions under which threats of sanctions promote cooperation and, when not, how they are likely to fail.

A large literature in psychology and economics points to both intentions and incentives as important determinants of sanctions' effects (Andreoni et al., 2003; Bewley, 1999; Camerer, 2003; Dickinson, 2001; Falk et al., 1999; Fehr and Falk, 2002; Fehr and Gächter, 2000a; Fehr and List, 2004; Fehr and Rockenbach, 2003; Fehr and Schmidt, 2007; Gneezy and Rustichini, 2000a,b; Ostrom et al., 1992; Sefton et al., 2002; Yamagishi, 1986, 1988). In general, intentions involve personalized rules by humans while incentives involve impersonal rules by law. In this paper, "incentive effects" refer narrowly to pecuniary tradeoffs a sanctioning mechanism creates while "intention" effects are those stemming from a punishment recipient's belief regarding a punisher's motivation for threatening (or not) a sanction. Unfortunately, to our knowledge, the respective roles of punishment intentions and incentives have not been previously distinguished. We present here an experiment designed to provide compelling systematic evidence on this issue. Our experiment distinguishes "types" of non-cooperative behavior, and we draw inferences regarding how non-cooperative decisions differ among various incentive and intention conditions.

Our design is a gift-exchange experiment (Camerer and Weigelt, 1988; Fehr et al., 1993; Berg, et al., 1995) where an investor sends an amount to a trustee, that amount is tripled by the experimenter, and then the trustee sends some fraction of the tripled amount back to the investor. Following Fehr and Rockenbach (2003) (henceforth F&R), we allow investors to specify their desired back-transfer amount, and decide whether to sanction trustees if the back-transfer is less than their request.

F&R report that trustees return less on average when investors use sanctions to enforce unfair outcomes. However, sanctions in their study only occur due to investors' intentional choices. Consequently, F&R's data cannot offer evidence on the separate roles of incentives and negative intentions in undermining cooperation. Our design compares outcomes between a replication of F&R's "intentions" treatment and a new treatment where sanction threats are randomly assigned to trustees by nature. In this way we provide cogent evidence distinguishing sanctions' intention and incentive effects.

Over 500 people participated in our experiment. We replicate the key patterns found in F&R's original data, and in particular punishment's detrimental effect. When not threatened, a trustee typically decides to return a positive amount less than the investor requested. When threatened this decision is least common. In particular, if the request is large relative to the sanction then the most common decision is to return nothing.¹ If the request is relatively small, then trustees generally return everything. Moreover, we find these patterns in both fair and unfair (unequal payoffs favoring the investor) backtransfer request cases. Finally, and importantly, we obtain compelling evidence that sanctions have statistically indistinguishable effects on trustees regardless of whether trustees are threatened intentionally by investors or randomly by nature.

Substantial evidence indicates intentions matter in many economic environments (Charness 2004; Charness and Levine, 2003; Charness and Haruvy, 2002; Falk et al., 2007, 2003; Fehr et al., 1993; McCabe et al., 2003). We discuss below the psychological concepts of cognitive dissonance (Akerlof and Dickens, 1982; Festinger, 1957; Konow, 2000; Rabin, 1994) and self-serving bias (Babcock and Loewenstein, 1997), and point out these theories can reconcile our and others' findings regarding intention effects.

2. Background

2.1. Incentive Effects

Standard economic arguments show incentives can enforce cooperation. Threats of sanctions can reduce expected payoffs to non-cooperative actions and leave cooperation

¹ Other studies report similar all-or-nothing behavior (Dickinson, 2001; Fehr and Gächter, 2002; Tyran and Feld, 2006). However, we are not aware of any study that distinguishes the relative importance of intention and incentive effects in these all-or-nothing decisions, nor have these studies recognized the significance of punishment's relative magnitude in affecting the distribution of returns.

relatively more attractive. On the other hand, many have pointed out that incentives can reduce cooperation (Fehr and Falk, 2002; Frey and Oberholzer-Gee, 1997; Kreps, 1997; Pokorny, forthcoming; Tenbrunsel and Messick, 1999). Gneezy and Rustichini (2000a), for example, found the number of late-coming parents increased when a small fine was imposed on parents who arrived late to collect their children. An explanation they suggested is that fines change parents' perception of extra-care service from a generous, non-market activity to a market commodity, and whether to obtain extra-care changes from a social norm to a price-based market decision.²

Reasoning such as appears in Gneezy and Rustichini (2000a) is quite familiar to psychologists. In particular, cognitive dissonance theory (Festinger, 1957) posits that people desire to keep their behavior and beliefs consistent. Absent external incentives, people justify their behavior through an appeal to internal motivations. However, when present, an external incentive can become a salient behavioral justification and this can, in principle, crowd out norm-based conduct (Deci et al., 1999; Lepper and Greene, 1978). This theory can be extended in an important way by an appeal to "self-serving bias", which argues individual's judgments are biased toward self-interest (Babcock and Loewenstein, 1997). Thus, in the case of fines, cognitive dissonance theory suggests threatened individuals might base their decisions on the external incentive and if so, due to self-serving bias, they are likely to interpret the fine as a price in order to support selfish behavior.

Thus, incentives are predicted to affect behavior in our environment as follows. First, people will cooperate to avoid punishment when a sanction's cost exceeds the benefit of defection. Second, when a sanction's cost is lower than this, people will pay the fine and selfishly maximize earnings. Finally, absent threats people will make decisions within an ethical context and selfish decisions will be relatively infrequent.

2. 2. Intention Effects

When one person chooses to threaten credibly another, both the incentives of the mechanism and also the intentions that underlie the threat can affect behavior. Humans

² Kreps (1997) posits screening and selection as sources of negative incentive effects. These aspects of incentives do not play a role in our environment.

are strongly disposed to infer intentionality when understanding others' actions (Gibbs, 1999; Kahneman et al., 1986; Taylor, 1979), and many economic models now include intention effects. Rabin (1993) is an early approach to incorporating the perceived kindness of another into one's own preferences. Another nice example is Dufwenberg and Kirchsteiger (2004), who develop a theory of reciprocity for extensive form games. A substantial amount of experimental research also suggests that intentions can play an important role in shaping decisions (Blount, 1995; Brandts and Solà, 2001; Charness, 2004; Charness and Haruvy, 2002; Charness and Levine, 2003; Falk et al., 2007, 2003; Falk and Kosfeld, 2004; Fehr and Gächter, 2000 b; Fehr et.al., 2007; Greenberg and Frisch, 1972; Gordon and Bowlby, 1989; McCabe et al., 2003; Nelson, 2002; Offerman, 2002).³

Imposing sanctions can be seen as a signal of distrust (Dickinson and Villeval, 2004; Fehr and Falk, 2002; Frey, 1993), or might create a hostile atmosphere (Bewley, 1999) and consequently reduce cooperation.⁴ In our experiment, an implication is that an investor who threatens a sanction, even one meant to enforce an earnings-equalizing backtransfer, might receive substantially less than when the same incentive is used without intentions (e.g., by assigning it randomly as in our experiment). Of course, intentions can also have a positive effect on cooperation: investors who choose not to sanction might be perceived as “nice” and relatively highly rewarded.

Our design provides transparent inference with respect to the way intentions influence reactions to threats, or non-threats, and is to our knowledge the first to separately identify effects of sanction incentives and intentions on non-cooperation.

3. Experiment Design

3.1. Fehr and Rockenbach's Design

Our experiment builds on F&R (2003), and we begin by discussing their experiments. F&R study sanction effects in a modified investment game (Berg et al., 1995). In this

³ Bolton, et al. (1998) report data from a design where intentions do not seem to affect behavior.

⁴ Charness and Dufwenberg (2006) suggest that investors who threaten sanctions might be perceived by trustees as expecting a low return, and that trustees would therefore feel less guilty about returning little. Also, Benabou and Tirole (2003) show performance incentives intentionally implemented by a principal who has more information about a task than an agent can adversely impact an agent's perception of a task or of her own abilities. The principal in our experiment does not have more information than the agent. Thus, our design cannot inform this potential intention effect.

game, both investor and trustee receive an endowment of ten monetary units (MUs). The investor sends some, all or none of his endowment to the trustee, and the experimenter triples any amount sent. In addition, the investor specifies an amount, between zero and the entire tripled amount, that she would like the trustee to send back. After observing the tripled amount and the desired back-transfer, the trustee sends some, all or none of the tripled amount to the investor. The investor earns his endowment of 10 MUs, minus the transfer amount, plus any back-transfer. The trustee earns the endowment of 10 MUs, plus the tripled transfer amount, less any back-transfer.

F&R introduce an “incentive” condition which is identical to the trust game baseline except the investor can now choose whether to commit to imposing a fine of four MUs on the trustee if less than the desired backtransfer amount is returned. Both subjects are aware that this fine, if due, does not accrue to the investor but rather to the experimenter’s research budget. Also, when the trustee makes her decision in the incentive condition she knows which sanction option her investor has chosen. F&R find that, on average, trustees’ back-transfers are lower when investors impose fines. Hence, F&R provide evidence that sanctions can reduce cooperation.⁵

3.2. Design

Our design allows us to compare conditions with and without punishment intentions, thus enabling inferences regarding punishment intention and incentive effects.

Figure A describes our intentions treatment, which replicates F&R’s “incentive” condition.⁶

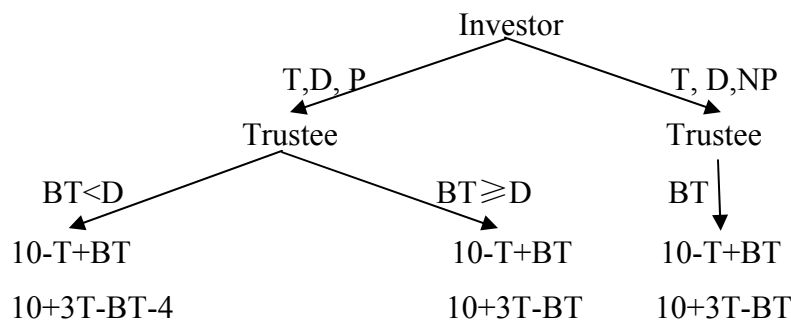


Fig. A. Intention treatment

⁵ Fehr and List (2004) report similar results using the same design with a non-student subject pool.

⁶ Figures A and B are not formal game trees. We use them only to assist in describing our experiment.

T—investor’s transfer; D—investor’s desired backtransfer; P—punishment threat; NP—no punishment threat; BT—trustee’s backtransfer.

Subjects are paired anonymously, and each subject is endowed with \$10. An investor chooses an integer amount in $[0,10]$ to transfer to her trustee (T), an integer amount in $[0, 3T]$ to request back (D), and whether to threaten punishment (P or NP). The trustee receives $3T$ and observes D and P or NP. Then, the trustee chooses an integer amount in $[0, 3T]$ to send to the investor (BT). If the investor chose to impose the payoff cut, then the earnings of the trustee are reduced by \$4 if the trustee returns less than requested backtransfer amount. The \$4 sanction does not accrue to the investor, but remains in the experimenter’s research budget. All of this is common knowledge.

Our second treatment, the “random” treatment, is described by Figure B.

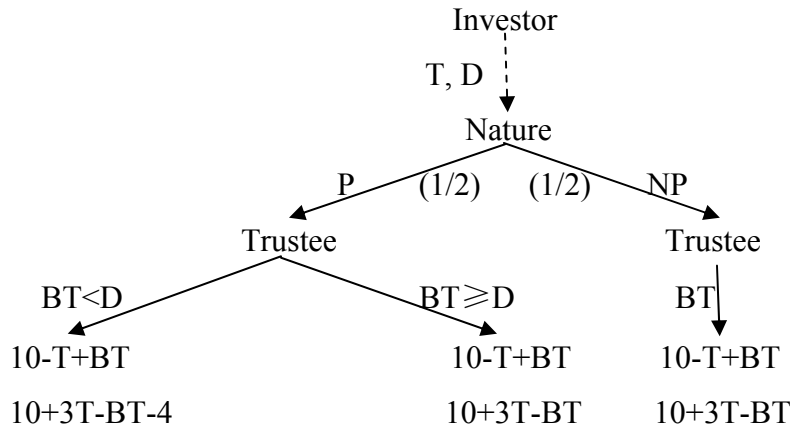


Fig. B. Random Treatment

Our goal with the random treatment is to eliminate any possibility that trustees could believe investors had punishment intentions. To do this we assign randomly the conditional \$4 payoff cut to exactly half of the trustees (indicated by the “(1/2)” in Figure B). Moreover, we give investors no information about sanctions that trustees might face (denoted by the dotted line in Figure B), and trustees know this is the case. It is clear to trustees that, at the time they make their decisions, investors only know the following: (i) investors choose the transfer amount, and trustees receive tripled that amount; (ii) investors request a desired backtransfer; and (iii) each trustee observes his/her investor’s decision before deciding the backtransfer amount (see Appendices A and B for transcripts of our instructions).

Blinding investors to the possibility that the trustee will be sanctioned is necessary for clean separation of intention and incentive effects. To see this, note that if an investor knows trustees might be sanctioned, then an investor can definitely avoid imposing a sanction threat if and only if she requests a backtransfer of zero. Hence, if an investor requests any positive backtransfer amount, then a trustee randomly assigned to the sanction condition might believe the investor “expected” or “hoped” the sanction would be assigned. It is plausible that trustees’ decisions could be affected by such intention attributions, thus confounding intention and incentive effects.

An obvious consequence of blinding investors to the sanctioning procedures is that, while they are of course not deceived, they also do not have full information about the way trustees’ payoffs are determined. This has the usual consequence that we lose control over what investors believe regarding trustees’ earnings and the way that might affect their own earnings. Accordingly, interpreting investor behavior is difficult. *Consequently, we draw no inferences regarding motivations for investor decisions. Our conclusions are based entirely on the decisions of trustees.*

It is worthwhile to note that the design of our random treatments provides no opportunity for trustees to exploit the information asymmetry. One reason is that the asymmetry is temporary: at the end of each session each investor is told about the sanctioning procedure, as well as the true condition to which her trustee was assigned. Moreover, at the beginning of each session of the random treatment, trustees are told that investors will be given this information. Hence, our design ensures the random treatment is maximally symmetric with our intention treatment. Trustees in the random treatment know that after they have made their decisions each investor will be made aware of the sanctioning procedure, and each investor will know whether her respective trustee was assigned to the sanction condition.

The only difference between our random and intention treatments is that investors cannot have punishment intentions in the former, and it is not reasonable to expect that trustees would assign punishment intentions to investors in the random treatment.

3.3. Procedures

A total of 532 subjects participated in the experiment: 149 and 117 pairs in the intention and random treatments, respectively⁷. All subjects were recruited from George Mason University's general student population, using procedures in place at the Interdisciplinary Center for Economic Science. Subjects earned a \$5 bonus for arriving to the lab on time. Subjects earned E\$ during the experiment, and at the end of the experiment the E\$ were exchanged for dollars at the rate of one-to-one.

Our specific procedures are detailed in the instructions found in Appendices A and B. Key aspects of our procedures include the following. Each treatment included a one-shot experiment, and subjects knew this. Investors were led to one room and trustees another. After they were separated each trustee was randomly and anonymously matched with one investor. Investors wrote their decisions on cards that were then delivered by the experimenter to the trustees. Trustees wrote their decisions on the same cards, and those decisions were delivered back to the investor by the experimenter. Earnings were calculated, subjects were paid privately and the experiment concluded. On average, subjects were in the lab for 90 minutes and earned about \$15 plus the show-up bonus.

It is important to emphasize our procedures for eliminating investors' punishment intentions. In the random treatment, investors were not informed that some trustees would be threatened with sanctions until the end of the experiment (after all decisions had been made), and trustees were aware this was the case. We assigned trustees to sanction conditions as follows. Prior to distributing investors' decision cards to trustees, the experimenter blindly selected exactly half of them (this treatment always included a multiple of four subjects) and assigned the sanction threat to each. Trustees observed the random draw, and saw clearly that investors were not connected to the randomization procedure. After this, the cards were distributed to trustees exactly as in the intention treatment.

⁷ Three investors in the random treatment transferred zero. Because trustees have no decision in these cases, we dropped these data from our analysis, leaving 114 observations in the random treatment.

3.4. Benchmark Theoretical Predictions

Benchmark theoretical predictions are based on standard conditions of self-interested earnings maximization. Absent sanction threats, self-interested trustees should return nothing, and investors should expect this and send nothing. However, this is not equilibrium in our intentions treatment, because there an investor can send one, request a backtransfer of three, and threaten a sanction. In this case a self-interested trustee will return three (to avoid losing four), so that the investor earns 12, and the trustee 10. Now, a sanction of four can enforce backtransfer requests of at most four: income-maximizing trustees facing larger requests will always return zero. Consequently, investors earn less than 12 if they send three or more. Investors can earn exactly 12 if they send two, request a backtransfer of four, and threaten sanctions. In this case the trustee earns 12, regardless of whether she sends four or zero. Therefore, the benchmark prediction for our intentions treatment is that investors send one (or two), request a backtransfer of three (or four), and threaten sanctions. In these cases, trustees are predicted to return the requested amount.

4. Results

4.1. Descriptive Statistics

Table 1 summarizes the decisions of our investors and trustees. The table includes only cases where the backtransfer request is “fair” (less than or equal to $2/3$ of the tripled investment amount.) Table 3, discussed below, describes behavior in the unfair request cases. We separate the data because the fair and unfair cases cannot be pooled⁸. For example, consider the percent of the tripled transfer amount returned by trustees. When the request is fair and a sanction is randomly imposed about 32% is returned, while in the same unfair case about 12% is returned, and the difference is significant ($p < 0.05$).⁹ Section 4.2 and 4.3 report the fair request data. We report unfair backtransfer requests in section 4.4.

Table 1 shows that 96 trustees in the random treatment received a fair backtransfer request, and 44 of these trustees were randomly assigned to the sanction

⁸ That trustees behave differently under unfair backtransfer requests could perhaps imply an unfair intention effect. However, neither we nor F&R have the appropriate contrast available to assess this possibility.

⁹ Unless otherwise noted, p -values reported below correspond to two-sided, two-sample t -tests, assuming unequal variances, of the null hypothesis that the means are the same.

condition. In the intention treatment, 122 trustees were asked to return a fair amount, of whom 62 faced a threat. In both cases trustees return 32% of the tripled investment amount when threatened with punishment, and just slightly less in the no-threat conditions, at 29% and 27% in the random and intention conditions, respectively.

Table 1 also includes summary statistics from F&R's intention condition, which our intention treatment replicates. F&R (2003) discuss their data at length, and we will not do so here. It is important to note, however, that neither we nor F&R find statistically significant effects of sanction threats on mean percentage tripled amounts returned in our respective intention treatments when the backtransfer request is fair ($p=0.33$ and $p=0.40$ in our and F&R's data, respectively).

Clearly though, the effect of a \$4 sanction threat might depend on the size of the backtransfer request. To investigate interactions between backtransfer requests and sanctions we analyze two cases: one where the desired backtransfer is less than eight, and the other where it is eight or more.¹⁰ We use eight because (i) a request of four implies cooperating yields the same earnings as returning zero; (ii) there are additional psychic costs to incurring punishment (Masclet, et. al., 2003); and (iii) a cut-off of eight leaves roughly equal numbers of observations in both the high and low request categories.¹¹ Our results are robust to perturbations in the cutoff value.¹²

4.2. Intentions, Incentives and Backtransfers

Each trustee receives three messages: a transfer amount, a desired back-transfer and whether a sanction has been threatened (either by the investor or randomly by nature). We model trustees' decisions as a function of those three messages and an error component. We assume the error is independent across subjects in different sessions but allow it to be correlated among subjects within the same session.

We begin by investigating determinants of the percentage of the tripled amount returned by trustees. The first column of Table 2 provides results from an OLS

¹⁰ F&R do not have enough data to analyze separately high and low request cells. Thus, we cannot compare our results to theirs in this regard.

¹¹ In the random treatment, 42% of trustees who do not face threats of sanctions are asked to return less than 8E\$, while that number is 45% for trustees assigned to the conditional payoff cut. In the intention treatment those same numbers are 42% and 47% for the no-threat and threat cases, respectively.

¹² Results using alternative cutoff values, as well as our data, are available on request.

regression¹³ (allowing correlated errors within sessions) of the percentage of tripled amount returned on nine regressors: the investment amount and eight terms determined by interacting decisions to punish and not punish, randomly or intentionally, with the high and low request conditions. Our results are as follows.

Result 1. Punishment incentives, but not intentions, affect mean percent returned.

We ran a joint test of the four restrictions that each intention coefficient is equal to its corresponding random coefficient. The null hypothesis is that the coefficient on intention threat is equal to the coefficient on random threat, and so on for the three remaining comparisons. The alternative is that the coefficients of at least one pair take different values. This test is unable to reject the null ($F(4,35)=0.21$, $p=0.93$), providing evidence that punishment threatened by investors does not affect trustees' return decisions differently than threats imposed randomly by nature.

On the other hand, a test for the effect of punishment (that is, random punishment equals random no punishment, and so on for each of the three remaining pairs) provides evidence that sanction threats do influence trustees' decisions ($F(4,35)=2.54$, $p=0.06$).

Figure 1 summarizes these findings by plotting the percentage of tripled amount returned against the amount sent for each of the eight conditions of interest (random threat high request, and so on.) The legend details the number of trustees observed in each of the eight cases. Mean decisions in the paired intention and random conditions are statistically indistinguishable, suggesting that intentions do not affect these decisions. Moreover, absent sanctions, the percentage returned does not vary with the transfer amount. On the other hand, there is clear visual evidence that sanctions increase returns when backtransfer requests are low, but work in the opposite direction when requests are high.

Result 2: The variance of the percentage-return distribution is higher under punishment incentives.

¹³ We report OLS results because they are easy to interpret. The results are not substantively changed by running a Tobit that accounts for the fact that returned amounts are censored at 0% and 100%.

In light of Result 1, we pool the random and intention data, so that the eight cells analyzed above collapse into four. When the desired backtransfer request is less than eight the distributions have significantly higher standard deviations when punishment is threatened than when it is not (30.4 vs. 22.7, $(F(48,46)=1.79, p=0.02, \text{one-tail})$). With high requests, under threats of punishment the standard deviation is 27.8, while it is 23.6 when punishment is not threatened ($(F(56,64)=1.38, p=0.10, \text{one-tail})$).

4.3. Sanctions and Non-Cooperation

In this section we provide an alternative perspective on our data that focuses on types of non-cooperation, which we measure with respect to investors' back-transfer requests.¹⁴ We define a trustee's choice set to include three mutually exclusive and exhaustive alternatives: cooperate completely (send at least as much as the investor requested); defect completely (send nothing); and weakly non-cooperate (send a positive amount less than the requested amount.)

We use a random utility model to analyze the resulting data. We assume that utility associated with each alternative depends on the same nine explanatory variables as used in section 4.2, as well as an alternative and subject specific error component which we allow to be correlated among individuals within the same session. We assume each subject chooses the alternative associated with his or her highest subjective utility, and an error structure that implies a multinomial logit specification characterizes choice behavior.

Result 3: Punishment incentives, but not intentions, affect non-cooperation.

The second and third columns of Table 2 detail the multinomial regression estimates supporting this, with weak non-cooperation as the baseline. The key findings are (i) investment amounts are statistically insignificant; (ii) a joint test of the null hypothesis that the random and intention coefficients are indistinguishable cannot be rejected ($\text{chi-square}(8)=6.1, p=0.63$); and (iii) a joint test of sanction effects (e.g., random

¹⁴ Our one-shot environment limits the amount of type classification that we can do. Cooperative types have been studied in dynamic environments by Houser and Kurzban (2002) and Kurzban and Houser (2005), among others. A robust procedure for type classification is provided by Houser et al. (2004).

threat = random no threat, and so on) indicates trustees' decisions to cooperate or defect are significantly affected by threats of sanctions ($\chi^2(8)=38.6, p<0.01$).

It is worth emphasizing that we find no evidence supporting punishment intention effects. The fractions of trustees who return nothing when threats are combined with high backtransfer requests are insignificantly different at 42% and 50% in the intention and random treatments, respectively ($p=0.58$). Also, 72% of trustees cooperate when intentionally threatened with sanctions to enforce low requests, as compared to 65% who do so under random threats. The difference is not significant ($p=0.59$).

Result 4: Punishment reduces weak non-cooperation. Moreover, sanction threats increase complete cooperation when backtransfer requests are low but increase complete defection when requests are high.

In light of our above findings, we again pool the intention and random treatments' data. Figure 2 shows, for each of the resulting four cells, the pooled fraction of subjects who cooperated, weakly non-cooperated and completely defected. When the backtransfer request is low, 57% of trustees weakly non-cooperate in the absence of sanction threats while only 8% do so when threat are used, and this difference is significant ($p<0.01$). Similarly, when the backtransfer request is high, 51% of trustees weakly non-cooperate absent threats while under threats the frequency is significantly lower at 25% ($p<0.01$).

Differences in weak non-cooperation underlie differences in the "threat" and "no threat" distributions seen in Figure 2. The distribution in both "no threat" cells has an inverted "U" shape, while in both "threat" cells the distribution is "U" shaped. Using Kolmogorov-Smirnov tests, we find the two "no threat" distributions are statistically indistinguishable ($p>0.99$), and the two threat distributions are significantly different ($p<0.01$). Moreover, there is a significant difference between the no threat and threat distributions in the low and high backtransfer cases ($p<0.01$ and $p<0.10$, respectively).

Figure 2 shows that, when sanction threats are combined with low requests, more trustees return at least the requested amount than when threats are not used (69% vs. 23%, $p<0.01$). We noted the significant change in weak non-cooperation between these cases,

but the frequencies of defection (a zero return) are nearly indistinguishable between the low request no-threat and threat conditions, at 19% and 22%, respectively ($p=0.69$).

On the other hand, when a high backtransfer request is combined with a sanction threat 46% of trustees return nothing, while only 25% do so if a high request is not accompanied by a threat. The difference is significant ($p=0.01$). We pointed out above that threats also have a significant effect on weak non-cooperation. However, the fraction of trustees who cooperate is not statistically different between the high request, threat and no threat cases, at 30% and 25%, respectively ($p=0.52$).

Our finding that sanctions' effects depend on the size of the backtransfer they are used to enforce is consistent with a tendency to maximize profits under incentives due to cognitive dissonance and self-serving bias (see section 2.1).

4.4. Unfair Backtransfer Requests

Turn now to cases where investors requested more than 2/3 of the tripled amount. We observed 45 investors to do this. Mean decisions made by investors and trustees in these unfair cases, again both in our experiments and F&R's, are detailed in Table 3. Unfair backtransfer requests are typically combined with a sanction threat. In the intention treatment we observed only four investors out of 27, and F&R four out of 15, who chose an unfair backtransfer request but did not threaten punishment. It turns out that 12 of 18 unfair requests in our random treatment were assigned to the sanction condition.

Result 5: Punishment intentions do not affect percentage amounts returned when the backtransfer request is unfair. However, regardless of intentions, percentages returned under sanction threats are significantly lower than when no threats are present.

Before detailing support for this result, note that in five of our unfair observations the investor sent one and requested all three back. All of these occurred in the "intention" treatment, and all five investors combined this request with the sanction threat. This is the only case where a trustee who does not completely cooperate loses money from their endowment of ten: a trustee who returns nothing earns only nine in the experiment. One of the five corresponding trustees in fact returned nothing, but the others cooperated by

sending back everything. Including these special cases, Table 3 shows mean percentages of tripled amounts returned in the intention and random treatments are 27% and 12%, respectively. However, this is not evidence that punishment intentions have a positive effect on cooperation, because we observe no such case in the random treatment.

We exclude the five special cases and compare means over our remaining 18 cases where punishment was intentional to the 12 cases where it was not (Table 3, columns 5 and 2, respectively). Entry by entry, the columns are nearly indistinguishable. Columns 1 and 3 of Table 3, the no-threat cases, are nearly point-wise identical. There are no significant differences between any pair of entries in columns 1 and 3, or 2 and 5.

In light of this, we pool the intention and random data. Excluding the five special cases discussed above (F&R report no such observations), we found trustees return 12% of the tripled transfer amount when fines are used to enforce unfair backtransfers. Absent fines twice this amount (25%) is returned, and the increase is significant (Mann-Whitney test, $p=0.03$, two-tail). This result, combined with our finding that there is no significant sanction effect on mean percentage returns in the case of fair requests (see section 4.1), replicates F&R's principal finding.

We also examine non-cooperative behavior. We have only four observations where the sanction is combined with a request of less than eight.¹⁵ Thus, we focus attention on cases where the unfair backtransfer request is eight or more. Absent sanction threats, 89% of these subjects chose to return a positive amount less than the investor requested. Under threats the number drops to 42%, and this change is statistically significant (Mann-Whitney test, $p=0.02$, two-tail). In addition, the most common decision under threats is to return nothing, occurring with frequency 54%. Thus, we find weak non-cooperation in our unfair data follows the pattern in our fair backtransfer data.

These findings can be summarized as follows. Our analysis, advantaged by our random treatment, suggests that incentives are the reason punishment has negative effects on mean returns when used to enforce unfair outcomes in this environment. We reported evidence that when backtransfer requests are high in relation to a sanction's size, regardless of whether the request is fair and regardless of whether punishment is

¹⁵ This is not unexpected, as F&R have only one observation on this case.

intentional, punishment incentives have detrimental effects on cooperation. In both our and F&R's unfair request data, when sanctions were used, it was almost always in an effort to enforce relatively high requests (10 of 11 investors in F&R's experiment request eight or more, and 26 of 30 investors did so in our own experiment.) This led to detrimental effects on mean returns in the unfair backtransfer data. On the other hand, in fair request cases sanction effects on mean returns are masked because threats are used to enforce high and low requests in nearly equal numbers: In F&R, 9 of 19 fair requests with sanctions included low requests, while our split is 49 of 106.

5. Conclusion

We reported data from a novel gift exchange experiment. Credible threats of sanctions often failed to produce cooperative behavior, and our evidence is that incentives, not intentions, underlie this effect. We found that trustees are significantly more likely to return nothing to investors when a relatively small sanction threat is used to enforce a relatively large backtransfer request. This is true regardless of whether the threat originated with the investor and irrespective of whether the request is fair.

Our results are consistent with previous studies that find negative incentive effects due to cognitive dissonance and self-serving bias (see Section 2.1 above.) The relevant implications of these theories are that sanction threats change individuals' perceptions of the environment, and can reduce internal motivations for cooperation (e.g., a preference to obey social norms). In our case these theories imply subjects interpret punishment as the price for self-interested behavior (Gneezy and Rustichini, 2000a) and the price, regardless of whether it was intentionally imposed, is an excuse for selfishness.¹⁶

Cognitive dissonance and self-serving bias can also explain why intentions do play a role in many other economic environments (Charness, 2004; Charness and Haruvy, 2002; Charness and Levine, 2003; Falk et al., 2003; Falk and Kosfeld, 2004). For example, McCabe et al. (2003) found, in a simplified two-person trust game, trustees are more likely to behave cooperatively when investors can make a decision that signals trust. This signal denies trustees the opportunity to interpret the game as one that does not

¹⁶ In post-experiment questionnaires many trustees who had not been threatened reported guilt-avoidance was an important reason to send money back, while threatened trustees often said they maximized their own earnings.

require reciprocity through trustworthy behavior. Similarly, when an unequal division in an ultimatum game is decided exogenously, so is not connected to signals of unfairness (Blount, 1995; Nelson, 2002), a responder can use this to fact to excuse their profit maximizing decision to accept the offer.

In the naturally occurring world, sanctions used to enforce pro-social behavior are often smaller than the cost of cooperation. Examples include fines for parking too long or not removing snow from sidewalks quickly enough. In light of our findings, one might question why such systems remain prevalent. A reason is that counterfactual analyses are difficult, so negative effects of such sanctions can be difficult to detect. Another is that sanctions create the politically expedient impression that “something is being done”. Finally, sanctions are often publicly enforced, and this can potentially mitigate their detrimental effects (Xiao and Houser, 2006).

We are extending this study to dynamic environments, and are also investigating links between intentions and non-monetary sanctions (Masclet et. al., 2003, Xiao and Houser, 2005). The role of punishment in economic exchange is an important topic of enduring interest. This paper is a step towards an improved understanding of when punishment works, and how it can fail.

References

- Akerlof, G.A., Dickens, W., 1982. The Economic Consequence of Cognitive Dissonance. *Amer. Econ. Rev.* 72(3), 307-319.
- Andreoni, J., Harbaugh, W., Vesterlund, L. 2003. The Carrot or Stick: Reward, Punishment and Cooperation. *Amer. Econ. Rev.* 93(3), 893-902.
- Babcock, L., Loewenstein, G. 1997. Explaining Bargaining Impasse: The Role of Self-Serving Biases. *J. Econ. Perspect.* 11(1), 109-126.
- Benabou, R., Tirole, J. 2003. Intrinsic and Extrinsic Motivation. *Rev. Econ. Stud.* 70, 489-520.
- Berg J., Dickhaut J., McCabe K. 1995. Trust, Reciprocity and Social History. *Games Econ. Behav.* 10, 122-142.
- Bewley, T. 1999. *Why Wages Don't Fall During a Recession*. Harvard University Press, Cambridge, MA.
- Blount, S. 1995. When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences. *Organizational Behavior and Human Decision Process* 63 (2), 131-144.
- Bolton, G., Brandts, J., Ockenfels, A. 1998. Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Game. *Exper. Econ.* 1, 207-219.
- Brandts, J., Solà, C. 2001. Reference Points and Negative Reciprocity in Simple Sequential Games. *Games Econ. Behav.* 36, 397-409.
- Camerer, C., Weigelt, K. 1988. Experimental tests of a sequential equilibrium reputation Model. *Econometrica.* 56, 1-36.
- Camerer, C. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.
- Charness, G., Haruvy, E. 2002. Altruism, Fairness, and Reciprocity in a Gift-Exchange Experiment: An Encompassing Approach", *Games Econ. Behav.* 40, 203-231.
- Charness, G., Levine, D. 2003. *The Road to Hell: An Experimental Study of Intentions*. Working paper.

- Charness, G., Dufwenberg, M. 2006. Promises and Partnership. *Econometrica*. 74(6), 1579-1601.
- Charness, G. 2004. Attribution and Reciprocity in an Experimental Labor Market. *J. Lab. Econ.* 22, 665-668.
- Deci, E.L., Koestner, R.M., Ryan, R. 1999. A meta-analytic review of experiments examining the effect of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 627–668.
- Dickinson, D. 2001. The Carrot vs. the Stick in Work Team Motivation. *Exper. Econ.* 4, 107-124.
- Dickinson, D., Villeval, M. 2004. Does Monitoring Decrease Work Effort? The Complementarity Between Agency and Crowding-Out Theories. *IZA Discussion Papers* 1222.
- Dufwenberg, M., Kirchsteiger, G. 2004. A theory of sequential reciprocity. *Games Econ. Behav.* 47, 268-98.
- Falk, A., Fehr, E., Fischbacher, U. 2007. Testing Theories of Fairness—Intentions Matter. *Games and Economic Behavior*. In press.
- Falk, A., Fehr, E., Fischbacher, U. 2003. On the Nature of Fair Behavior. *Econ. Inquiry*. 41(1), 20-26.
- Falk, A., Gächter, S., Kovács, J. 1999. Intrinsic Motivation and Extrinsic Incentives in a Repeated Game with Incomplete Contracts. *J. Econ. Psych.* 20(3), 251-284.
- Falk, A., Kosefeld, M. 2004. Distrust—The hidden Cost of Control. *IZA discussion paper* 1203.
- Fehr, E., Gächter, S. 2000 a. Cooperation and Punishment in Public Goods Experiments. *Amer. Econ. Rev.* 90 (4), 980-994.
- Fehr, E., Schmidt, K.M., 2007. Adding a Stick to the Carrot? The Interaction of Bonuses and Fines. *Amer. Econ. Rev. Papers and Proceedings*. 97,177-181.
- Fehr, E., Klein, A., Schmidt, K., 2007. Fairness and Contract Design. *Econometrica* 75, 121 154.
- Fehr, E., Gächter, S. 2000 b. Fairness and Retaliation: The Economics of Reciprocity. *J. Econ. Perspect.* 14(3), 159-181.

- Fehr, E., Gächter, S. 2002. Do Incentive Contracts Undermine Voluntary Cooperation. Working Paper, University of Zurich.
- Fehr, E., Rockenbach, B. 2003. Detrimental Effects of Sanctions on Human Altruism. *Nature*, 422, 137-140.
- Fehr, E., List, J. 2004. The Hidden Costs and Rewards of Incentives. *Journal of the European Economic Association*, 2(5), 743-771.
- Fehr, E., Falk, A. 2002. Psychological Foundation of Incentives. *Europ. Econ. Rev.* 46, 687 - 724.
- Fehr, E., Kirchsteiger, G., Riedl, A. 1993. Does fairness prevent market clearing? An experimental investigation. *Quart. J. Econ.* 108, 437-460.
- Festinger, L. 1957. *A theory of cognitive dissonance*. Stanford University Press, Stanford, CA.
- Frey, B., Oberholzer-Gee, F. 1997. The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding –Out. *Amer. Econ. Rev.* 87, 746-755.
- Frey, B. 1993. Does Monitoring Increase Work Effort? The Rivalry between Trust and Loyalty. *Econ. Inquiry*. 31, 663-670.
- Gibbs, R. 1999. *Intentions In The Experience Of Meaning*. Cambridge University Press. New York, NY, pp. 69-106,
- Gneezy, U., Rustichini, A. 2000a. A Fine Is A Price. *J. Legal Stud.* 29(1), 1-17.
- Gneezy, U., Rustichini, A. 2000b. Pay Enough or Don't Pay at All. *Quart. J. Econ.* 115(2), 791-810.
- Gordon, M., Bowlby, R. 1989. Reactance and Intentionality Attribution as Determinants of the Intent to File a Grievance. *Personal Psychology*, 42, 309-329.
- Greenberg, M., Frisch, D. 1972. Effect of Intentionality on Willingness to Return a Favor. *Journal of Experimental Social Psychology*, 8, 99-111.

- Houser, D., Keane, M., McCabe, K. 2004. Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence Using a Bayesian Type Classification Algorithm. *Econometrica*. 72(3), 781-822.
- Houser, D., Kurzban, R. 2002. Revisiting Kindness and Confusion in Public Goods Experiments. *Amer. Econ. Rev.* 92(4), 1062-1069.
- Kahneman, D., Knetsch, J., Thaler R. 1986. Fairness as a Constraint on Profit-Seeking: Entitlements in the Market. *Amer. Econ. Rev.* 76, 728-741.
- Konow, J. 2000. Fair Share: Accountability and Cognitive Dissonance in Allocation Decisions. *Amer. Econ. Rev.* 90(4), 1072-1091.
- Kreps, D. 1997. Intrinsic Motivation and Extrinsic Incentives. *Amer. Econ. Rev. Papers and Proceedings*. 87(2), 359-364.
- Kurzban, R., Houser, D. 2005. An Experiments Investigating Cooperative Types in Humans: A Complement to Evolutionary Theory and Simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20), 7398-7401.
- Lepper, M., Greene, D. 1978. *The Hidden Cost of Reward: New Perspectives on the Psychology of Human Motivation*, New York: John Wiley.
- Masclot, D., Noussair, C., Tucker, S., Villeval, M. 2003. Monetary and Non-monetary Punishment in the Voluntary Contributions Mechanism. *Amer. Econ. Rev.* 93, 366-380.
- McCabe, K., Rigdon, M., Smith, V. 2003. Positive reciprocity and intention in trust games. *J. Econ. Behav. Organ.* 52(2), 267-275.
- Nelson W.R. Jr., 2002. Equity and intention: it's the thought that counts. *J. Econ. Behav. Organ.* 48(4), 423-430.
- Ostrom, E., Walker, J., Gardner, R. 1992. Covenants With and Without a Sword: Self-Governance is Possible. *Amer. Polit. Sci. Rev.* 86, 404 - 417.
- Offerman, T. 2002. Hurting Hurts More Than Helping Helps. *Europ. Econ. Rev.* 46, 1423-1437.

- Pokorny, K. (forthcoming). Pay – But Don't Pay Too Much: An Experimental Study on the Impact of Incentives", *J. Econ. Behav. Organ.*
- Rabin, M. 1993. Incorporating Fairness into Game Theory and Economics. *Amer. Econ. Rev.* 83 (5), 1281-1302.
- Rabin, M. 1994. Cognitive Dissonance and Social Change *J. Econ. Behav. Organ.* 23(2), 177-194.
- Sefton, M., Shupp, R., Walker, J. 2002. The Effects of Rewards and Sanctions in Provision of Public Goods. Working paper.
- Taylor, C. 1979. Action as Expression, In: Diamond, C., Teichman, J. (Eds.), *Intention and Intentionality: Essays in Honor of G. E. M. Anscombe*. Cornell University Press, Ithaca, NY, pp. 73-89.
- Tenbrunsel, A. E., Messick, D. M. 1999. Sanctioning Systems, Decision Frames, and Cooperation. *Administrative Science Quarterly.* 44, 684-707.
- Tyran, J.R., Feld L. P. 2006. Achieving Compliance when Legal Sanctions are Non Deterrent. *Scandinavian Journal of Economics*, 108(1), 135-156.
- Xiao, E., Houser, D. 2005. Emotion Expression in Human Punishment Behavior. *Proceedings of the National Academy of Sciences*, 102(20), 7398-7401.
- Xiao, E., Houser, D. 2006. Punish in Public. IZA discussion paper 1977.
- Yamagishi, T. 1986. The provision of a Sanctioning System as a Public Good. *Journal of Personality and Social Psychology.* 51(1), 110-116.
- Yamagishi, T. 1988. Seriousness of Social Dilemmas and the Provision of a Sanctioning System. *Social Psychology Quarterly.* 51(1), 32-42.

Table 1
 Mean Decisions by Investors and Trustees when Request is Fair

	Random Treatment		Intention Treatment		Intention Treatment F&R (2003)**	
	No Threat	Threat	No Threat	Threat	No Threat	Threat
Investment	6.2 (0.4)	5.5 (0.4)	6.2 (0.4)	5.3 (0.4)	8.5 (0.7)	6.7 (0.7)
Desired back-transfer as a percentage of tripled investment	57.0 (1.7)	60.2 (1.3)	55.3 (1.5)	59.0 (1.2)	52.4 (6.4)	51.2 (4.1)
Actual back-transfer	5.7 (0.8)	4.3 (0.8)	5.2 (0.7)	4.2 (0.6)	11.0 (1.6)	7.3 (1.7)
Actual back-transfer as a percentage of tripled investment	29.0 (3.2)	31.8 (4.9)	27.2 (3.0)	31.9 (3.8)	43.0 (5.2)	35.2 (6.4)
Actual back-transfer as a percentage of request	53.4 (6.0)	53.1 (7.9)	54.5 (8.0)	56.3 (6.7)	78.1* (9.1)	71.8 (10.4)
Number of pairs	52	44	60	62	11	19

Note: Numbers in parentheses are standard errors.

* This mean excludes one pair whose investor requested a zero back-transfer.

** Fehr & Rockenbach (2003) label this the “incentive” condition.

Table 2
 Effect of Intentions and Incentives on (1) Percentage of Tripled Investment Amount Returned by Trustees and (2) Cooperation and Non-Cooperation

Variables	(1) OLS	(2) Multinomial Logit**	
	Percentage of Tripled Investment Amount Returned	Completely Defect (Return=0)	Completely Cooperate (Return≥Request)
	Coefficient	Coefficient	Coefficient
Investment	-0.38 (0.94)	-0.02 (0.07)	0.04 (0.08)
Intention threat	23.68 (7.58)	0.38 (0.57)	-0.63 (0.77)
Random threat	26.39 (11.01)	1.53 (0.65)	0.77 (1.14)
Intention no threat	30.65 (7.60)	-0.44 (0.67)	-1.27 (0.78)
Random no threat	34.72 (8.66)	-0.76 (0.68)	-0.85 (0.77)
Intention threat and low request*	21.88 (8.26)	0.19 (0.77)	2.44 (0.92)
Random threat and low request	16.50 (12.38)	0.31 (1.27)	1.67 (1.37)
Intention no threat and low request	-2.62 (6.67)	-0.44 (0.62)	0.48 (0.81)
Random no threat and low request	-7.87 (7.61)	-0.42 (0.68)	-0.56 (0.85)
	R ² =0.588	Pseudo R ² =0.128	

Note: Numbers in parentheses are robust standard errors.

*Low request means the desired back transfer is less than eight.

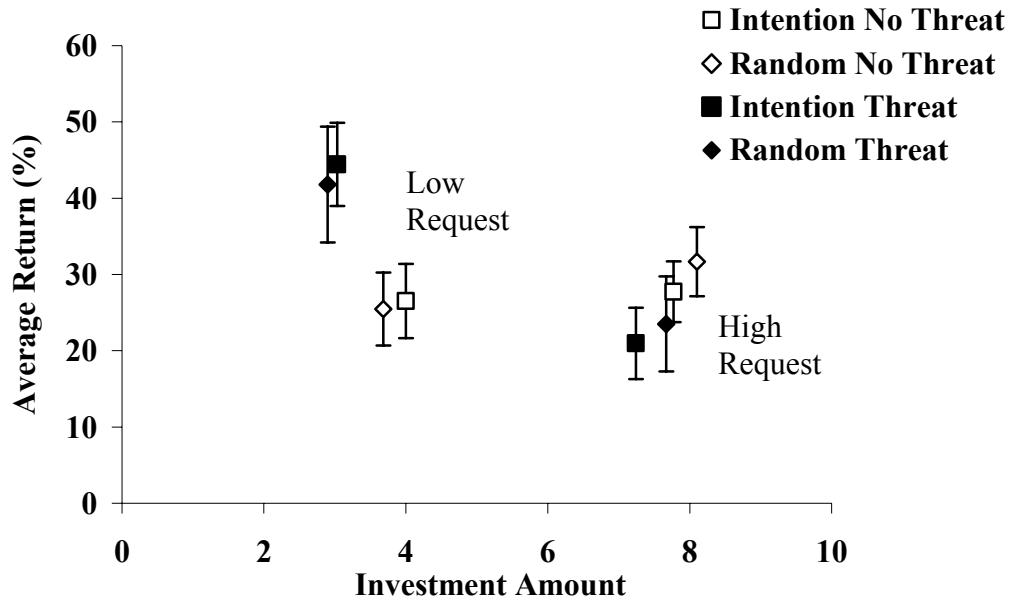
** The baseline is weak non-cooperation (0 < return < request).

Table 3
Mean Decisions by Investors and Trustees when Request is Unfair

	Random Treatment		Intention Treatment			Intention Treatment F&R (2003)**	
	No Threat	Threat	No Threat	Threat		No Threat	Threat
				All	Exclude (1,3)*		
Investment	5.3 (1.3)	5.8 (0.8)	7.0 (1.1)	5.2 (0.7)	6.3 (0.6)	9.0 (0.7)	7.1 (0.8)
Desired back-transfer as a percentage of tripled investment	91.2 (5.6)	83.2 (3.2)	91.9 (6.6)	92.3 (2.0)	90.1 (2.3)	94.6 (3.9)	95.4 (2.5)
Actual back-transfer	5.3 (2.5)	2.5 (1.3)	5.0 (1.5)	2.2 (0.7)	2.2 (0.9)	16.5 (2.4)	3.8 (1.6)
Actual back-transfer as a percentage of tripled investment	23.4 (10.0)	12.1 (6.7)	27.0 (8.0)	26.5 (8.1)	11.6 (4.7)	60.1 (4.5)	22.0 (10.0)
Actual back-transfer as a percentage of request	29.0 (13.2)	14.4 (8.4)	30.7 (9.5)	28.3 (8.3)	13.9 (5.7)	64.0 (6.3)	24.5 (10.9)
Number of pairs	6	12	4	23	18	4	11

*This column excludes five cases where the investor transferred one, requested three back and imposed the sanction of four. Of these, one trustee returned zero and four returned three. F&R observe no such cases.

** Fehr & Rockenbach (2003) labels this the “incentive” condition.



Let L denote “Low Request,” H denote “High Request,” N=“No threat,” T=“Threat,” R=“Random” and I=“Intention.” Then the number of trustee observations in each cell is as follows. LRN=22, LIN=25, LRT=20, LIT=29, HRN=30, HIN=35, HRT=24, HIT=33.

Fig. 1. Return Percentage of Tripled Investment Amount (Intention vs. Random)

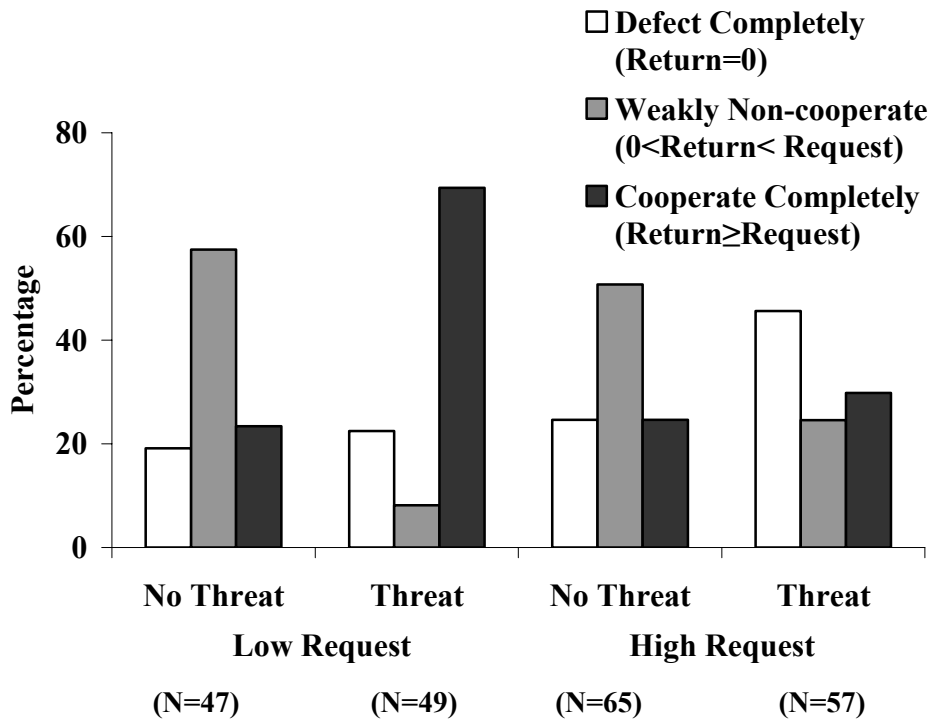


Fig. 2. Distribution of Trustees' Decisions (Fair Backtransfer Requests)

Appendix A: Instructions for the Intention Treatment

I. Investor's instruction

You are Actor 1

Description of Your Decision Problem

Thank you for coming! You've earned \$5 for showing up on time, and the instructions explain how you can make decisions and earn more money. So please read these instructions carefully! There is no talking at any time during this experiment. If you have a question please raise your hand, and an experimenter will assist you.

You are in Room A and you will be randomly paired with someone in Room B. You will never be informed of the identity of this person, either during or after the experiment. Similarly, your matched participant will never be informed about your identity. You are in the role of **Actor 1** and your matched participant is in the role of Actor 2. You and Actor 2 will participate only once in this decision problem. You make your decisions with the help of the decision sheet described below.

This is how the experiment works.

Endowment

Besides the \$5 show-up bonus, at the beginning of the experiment **both** actors receive an initial endowment of 10 E\$ (experimental dollars).

Your decision

Your decision includes three parts:

1. A transfer between 0 and 10 E\$ to Actor 2.

You, as Actor 1, can transfer, from your endowment, any amount between 0 and 10 E\$ to Actor 2. The experimenters will **triple** this transferred amount, so that Actor 2 receives three times the amount of E\$ you transferred.

2. A desired back-transfer.

You also make a decision about your desired back-transfer, that is, at least how many E\$ you would like to receive back from Actor 2. You can ask for any amount between zero and the tripled amount of your transfer.

3. Whether to impose a conditional payoff cut of 4 E\$ on Actor 2's final earnings.

- A conditional payoff cut of 4 E\$ for Actor 2 has the following consequences. The payoff of Actor 2 will be reduced by 4 E\$ if his/her actual back-transfer is less than your desired back-transfer. The conditional payoff cut does not happen if Actor 2 transfers your desired amount or more to you.
- If you choose not to impose a conditional payoff cut, then the income of Actor 2 will not be reduced, irrespective of the amount of Actor 2's back-transfer.

The decision of Actor 2

After your decision, Actor 2 can transfer back to you any amount of the tripled number of E\$ bills he/she received. In case that you have chosen a conditional payoff cut of 4 E\$, and if Actor 2 transfers back less than what you desired, then Actor 2 must pay the conditional cut.

Payoffs

You (Actor 1) receive: 10 E\$ – transfer to Actor 2 + back-transfer from Actor 2.

Actor 2 receives: 10 E\$ + 3 × transfer from Actor 1 – back-transfer to Actor 1 – 4 E\$ (in case that a conditional payoff cut was imposed and is due)

Exchange rate: For every E\$ you earn you will be paid \$1.

How the experiment is conducted

There are several envelopes in Room A and Room B. In each envelope in Room A and Room B, there is a card marked with a unique letter. Each envelope looks the same. Everyone in both Room A and Room B will randomly pick up an envelope. The person in Room B who chooses the card with the same letter as yours will be your Actor 2.

Items on your table: 10 E\$ bills (your endowment), two decision sheets (one for Actor 1 and one for Actor 2) and two **Yes/No** stickers.

Items on Actor 2's table: 10 E\$ bills (Actor 2's endowment)

You will make your decisions at your seat by filling in the decision sheets. You need to leave the number of E\$ bills you want to transfer in the envelope, but keep the card, which will help the experimenter to return the envelope to you later. Raise your hand after you're done. The experimenter will go to your seat, check whether all necessary information is on the decision sheets, then triple your transferred E\$ bills. The experimenter will also record the letter of your card on the back of the decision sheets so that your envelope can be given to your Actor 2 who has the card with the same letter. The experimenter will then put the decision sheets and tripled amount of E\$ bills in your envelope and collect it. After every Actor 1 has finished, the experimenter will take all the envelopes to Room B.

The experimenter will give each Actor 1's envelope to his/her Actor 2. Each Actor 2 then decides how much to transfer back to you by writing down a number on the decision sheets and leave the E\$ bills he/she wants to transfer back to you in the envelope. When Actor 2 has finished his/her decision, the experimenter will go to his/her seat, check whether all the necessary information is on the decision sheets and then put your copy of the decision sheet in the envelope. Actor 2 will keep his/her copy of the decision sheet.

After all Actor 2s have finished, the experimenter will take all the envelopes to Room A. The experimenter will return to you the envelope with the back-transfer E\$ bills from Actor 2 inside. Each Actor 1 will then be called, one by one, to the experimenter. When called, you will take your envelope to the experimenter. The experimenter will calculate your final earnings and pay you privately. Then please exit the lab. Since you will be asked to leave when you are done, you should take all your belongings when you go to the experimenter. Actor 2s in room B will be paid after **all** Actor 1s have been paid and have left the lab.

Throughout this experiment, you won't meet any Actor 2 in room B.

End of Instructions

Please raise your hand to indicate that you are finished reading these instructions. When you do, an experimenter will give you a short quiz to ensure that you understand how you make decisions.

II. Trustee's instructions

You are Actor 2

Description of Your Decision Problem

Thank you for coming! You've earned \$5 for showing up on time, and the instructions explain how you can make decisions and earn more money. So please read these instructions carefully! There is no talking at any time during this experiment. If you have a question please raise your hand, and an experimenter will assist you.

You are in Room B and you will be randomly paired with someone in Room A. You will never be informed of the identity of this person, either during or after the experiment. Similarly, your matched participant will never be informed about your identity. You are in the role of **Actor 2** and your matched participant is in the role of Actor 1. You and Actor 1 will participate only once in this decision problem. You make your decisions with the help of the decision sheet described below.

This is how the experiment works.

Endowment

Besides the \$5 show up bonus, at the beginning of the experiment **both** actors receive an initial endowment of 10 E\$ (experimental dollars).

The decision of Actor 1 (You are not Actor 1)

First Actor 1 has to make a decision that consists of the following three components.

1. A transfer between 0 and 10 E\$ to you.

Actor 1 can transfer, from his/her endowment, any amount between 0 and 10 E\$ to you. The experimenters will **triple** this transferred amount, so that you receive three times the amount of E\$ transferred by Actor 1.

2. A desired back-transfer.

Actor 1 will indicate his/her desired back-transfer, which is at least how many E\$ he/she would like to receive back from you. Actor 1 can ask for any amount between zero and the tripled amount of his/her transfer.

3. Whether to impose a conditional payoff cut of 4 E\$ on your final earnings.

- A conditional payoff cut of 4 E\$ has the following consequences for you. Your payoff will be reduced by 4 E\$ if your actual back-transfer is less than the back-transfer desired by Actor 1. The conditional payoff cut does not happen if you transfer the desired amount or more to Actor 1.
- If Actor 1 chooses not to impose a conditional payoff cut, then your income will not be reduced, irrespective of the amount of your back-transfer to Actor 1.

Your decision

After Actor 1 makes his/her decisions, you, as Actor 2, can transfer back to Actor 1 any amount of the tripled number of E\$ bills you received. As noted, in case that Actor 1 has chosen a conditional payoff cut of 4 E\$, and you transfer back less than what he/she desired, then you must pay the conditional cut.

Payoffs

Actor 1 receives: 10 E\$ – transfer to Actor 2 + back-transfer from Actor 2.

You (Actor 2) receive: 10 E\$ + 3× transfer from Actor 1 – back-transfer to Actor 1 – 4 E\$ (in case that a conditional payoff cut was imposed and is due)

Exchange rate: For every E\$ you earn you will be paid \$1.

How the experiment is conducted

There are several envelopes in Room A and Room B. In each envelope in Room A and Room B, there is a card marked with a unique letter. Each envelope looks the same. Everyone in both Room A and Room B will randomly pick up an envelope. The person in Room A who chooses the card with the same letter as yours will be your Actor 1.

Items on your table: 10 E\$ bills (Your endowment).

Items on Actor 1’s table: 10 E\$ bills (Actor 1’s endowment), two decision sheets (one for Actor 1 and one for Actor 2, as shown below) and two **Yes/No** stickers.

Sample Decision Sheets

Decision Sheet	Copy for Actor 1
Actor 1:	
1. I decide to transfer _____ E\$ to Actor 2	
2. My desired back-transfer amount: _____ E\$	
3. If Actor 2’s back-transfer is less than my desired back-transfer amount, I will impose a conditional payoff cut of 4 E\$ on Actor 2:	
Yes	No
Actor 2:	
Based on Actor 1’s decision, I decide to transfer _____ E\$ back to Actor 1	

Decision Sheet	Copy for Actor 2
Actor 1:	
1. I decide to transfer _____ E\$ to Actor 2	
2. My desired back-transfer amount: _____ E\$	
3. If Actor 2’s back-transfer is less than my desired back-transfer amount, I will impose a conditional payoff cut of 4 E\$ on Actor 2:	
Yes	No
Actor 2:	
Based on Actor 1’s decision, I decide to transfer _____ E\$ back to Actor 1	

Actor 1 will make his/her decision at his/her seat by filling in the decision sheets. Actor 1 will leave the number of E\$ bills he/she wants to transfer in the envelope he/she picked up, but keep the card, which will help the experimenter to return the envelope to him/her later. When Actor 1 is done, the experimenter will go to his/her seat, check whether all necessary information is on the decision sheets, then triple the transferred E\$ bills. The experimenter will also record the letter of his/her card on the back of the decision sheets so that his/her envelope can be given to his/her Actor 2 who has the card with the same letter. The experimenter will then put the decision sheets and tripled number of E\$ bills in his/her envelope and collect it. After every Actor 1 has finished, the experimenter will take all the envelopes to Room B.

The experimenter will give each of you your Actor 1’s envelope. When you get the envelope, decide how much to transfer back to Actor 1 by writing down a number on the decision sheets and leave the E\$ bills you want to transfer back to Actor 1 in the envelope. Raise your hand when you are done. The experimenter will go to your seat, check whether all the necessary information is on the decision sheets and then put Actor 1’s copy of the decision sheet in the envelope and collect it. You will keep your copy of the decision sheet. Don’t show anybody else your decision sheet.

After all Actor 2s have finished, the experimenter will take all the envelopes to Room A. The experimenter will return to Actor 1 his/her envelope with the back-transfer E\$ bills from Actor 2 inside. Each Actor 1 will be called, one by one, to the experimenter. The experimenter will calculate his/her final earnings and pay him/her privately. Then Actor 1 will exit the lab. After all Actor 1s have left, the experimenter will call each Actor 2 one by one. When called, you will go to the experimenter with your decision sheet. The experimenter will calculate your earnings and pay you privately.

Throughout this experiment, you won't meet any Actor 2 in room B.

End of Instructions

Please raise your hand to indicate that you are finished reading these instructions. When you're done, an experimenter will give you a short quiz to ensure that you understand how you make decisions.

Appendix B: Instructions for the Random Treatment

I. Investor's instruction

You are Actor 1

Description of Your Decision Problem

Thank you for coming! You've earned \$5 for showing up on time, and the instructions explain how you can make decisions and earn more money. So please read these instructions carefully! There is no talking at any time during this experiment. If you have a question please raise your hand, and an experimenter will assist you.

You are in Room A and you will be randomly paired with someone in Room B. You will never be informed of the identity of this person, either during or after the experiment. Similarly, your matched participant will never be informed about your identity. You are in the role of **Actor 1** and your matched participant is in the role of Actor 2. You and Actor 2 will participate only once in this decision problem. You make your decisions with the help of the decision sheet described below.

This is how the experiment works.

Endowment

Besides the \$5 show-up bonus, at the beginning of the experiment **both** actors receive an initial endowment of 10 E\$ (experimental dollars).

Your decision

Your decision includes two parts:

1. A transfer between 0 and 10 E\$ to Actor 2.

You, as Actor 1, can transfer, from your endowment, any amount between 0 and 10 E\$ to Actor 2. The experimenters will **triple** this transferred amount, so that Actor 2 receives three times the amount of E\$ you transferred.

2. A desired back-transfer.

You also make a decision about your desired back-transfer, that is, at least how many E\$ you would like to receive back from Actor 2. You can ask for any amount between zero and the tripled amount of your transfer.

The decision of Actor 2

After your decision, Actor 2 can transfer back to you any amount of the tripled number of E\$ bills he/she received.

Payoffs

You (Actor 1) receive: 10 E\$ – transfer to Actor 2 + back-transfer from Actor 2.

Exchange rate: For every E\$ you earn you will be paid \$1.

How the experiment is conducted

There are several envelopes in Room A and Room B. In each envelope in Room A and Room B, there is a tag marked with a unique letter. Each envelope looks the same. Everyone in both Room

A and Room B will randomly pick up an envelope. The person in Room B who chooses the tag with the same letter as yours will be your Actor 2.

Items on your table: 10 E\$ bills (your endowment), two decision sheets (one for Actor 1 and one for Actor 2).

Items on Actor 2's table: 10 E\$ bills (Actor 2's endowment)

You will make your decisions at your seat by filling in the decision sheets. You need to leave the number of E\$ bills you want to transfer in the envelope, but keep the tag, which will help the experimenter to return the envelope to you later. Raise your hand after you're done. The experimenter will go to your seat, check whether all necessary information is on the decision sheets, then triple your transferred E\$ bills. The experimenter will also record the letter of your tag on the back of the decision sheets so that your envelope can be given to your Actor 2 who has the tag with the same letter. The experimenter will then put the decision sheets and tripled amount of E\$ bills in your envelope and collect it. After every Actor 1 has finished, the experimenter will take all the envelopes to Room B.

Each Actor 2 will get his/her Actor 1's envelope according to the letter on their tags. Each Actor 2 then decides how much to transfer back to you by writing down a number on the decision sheets and leave the E\$ bills he/she wants to transfer back to you in the envelope. When Actor 2 has finished his/her decision, the experimenter will go to his/her seat, check whether all the necessary information is on the decision sheets and then put your copy of the decision sheet in the envelope. Actor 2 will keep his/her copy of the decision sheet.

After all Actor 2s have finished, the experimenter will take all the envelopes to Room A. The experimenter will return to you the envelope with the back-transfer E\$ bills from Actor 2 inside. Each Actor 1 will then be called, one by one, to the experimenter. When called, you will take your envelope to the experimenter. The experimenter will calculate your final earnings and pay you privately. Then please exit the lab. Since you will be asked to leave when you are done, you should take all your belongings when you go to the experimenter. Actor 2s in room B will be paid after **all** Actor 1s have been paid and have left the lab.

Throughout this experiment, you won't meet any Actor 2 in room B.

End of Instructions

Please raise your hand to indicate that you are finished reading these instructions. When you do, an experimenter will give you a short quiz to ensure that you understand how you make decisions.

II. Trustee's instructions

You are Actor 2

Description of Your Decision Problem

Thank you for coming! You've earned \$5 for showing up on time, and the instructions explain how you can make decisions and earn more money. So please read these instructions carefully! There is no talking at any time during this experiment. If you have a question please raise your hand, and an experimenter will assist you.

You are in Room B and you will be randomly paired with someone in Room A. You will never be informed of the identity of this person, either during or after the experiment. Similarly, your matched participant will never be informed about your identity. You are in the role of **Actor 2** and your matched participant is in the role of Actor 1. You and Actor 1 will participate only once in this decision problem. You make your decisions with the help of the decision sheet described below.

This is how the experiment works.

Endowment

Besides the \$5 show up bonus, at the beginning of the experiment **both** actors receive an initial endowment of 10 E\$ (experimental dollars).

The decision of Actor 1 (You are not Actor 1)

First Actor 1 has to make a decision that consists of the following two components.

1. A transfer between 0 and 10 E\$ to you.

Actor 1 can transfer, from his/her endowment, any amount between 0 and 10 E\$ to you. The experimenters will **triple** this transferred amount, so that you receive three times the amount of E\$ transferred by Actor 1.

2. A desired back-transfer.

Actor 1 will indicate his/her desired back-transfer, which is at least how many E\$ he/she would like to receive back from you. Actor 1 can ask for any amount between zero and the tripled amount of his/her transfer.

The randomly determined Conditional Payoff-Cut

Half of the Actor 2s will be randomly assigned to receive the Payoff-Cut and **half** randomly assigned not to receive the Payoff-Cut.

- If you are randomly assigned to the Payoff-Cut, there will be a conditional payoff cut of 4 E\$ for you. A conditional payoff cut has the following consequences: Your payoff will be reduced by 4 E\$ if your actual back-transfer is less than Actor 1's desired back-transfer. The conditional payoff cut does not happen if you transfer the desired amount or more to Actor 1.
- If you are randomly assigned to No Payoff-Cut, then your earnings will not be reduced, irrespective of the amount of your back-transfer.

Important:

- (1) **Actor 1s have not been told about the conditional payoff cut. When Actor 1s make their transfer and desired back-transfer decisions, they don't know that some Actor 2s will be assigned to a conditional payoff-cut.**
- (2) **Whether you are assigned to the Payoff-Cut is randomly determined. Nothing that you or Actor 1 does affects whether you are assigned to the Payoff Cut. Each Actor 2's assignment is indicated by the Payoff-Cut sticker and No Payoff-Cut sticker (as explained below).**
- (3) **At the end of the experiment, *after all decisions have been made*, Actor 1s will be informed about the Payoff-Cut, and will be told whether their matched Actor 2s were randomly assigned to the Payoff-Cut.**

Your decision

After Actor 1 makes his/her decisions, and after whether you are assigned to the conditional payoff-cut has been randomly determined, you, as Actor 2, can transfer back to Actor 1 any amount of the tripled number of E\$ bills you received. As noted, in case that you are assigned to the conditional payoff cut of 4 E\$, and you transfer back less than what he/she desired, then you must pay the conditional cut.

Payoffs

Actor 1 receives: 10 E\$ – transfer to Actor 2 + back-transfer from Actor 2.

You (Actor 2) receive: 10 E\$ + 3× transfer from Actor 1 – back-transfer to Actor 1 – 4 E\$ (in case that a conditional payoff-cut is due).

Exchange rate: For every E\$ you earn you will be paid \$1.

How the experiment is conducted

There are several envelopes in Room A and Room B. In each envelope in Room A and Room B, there is a tag marked with a unique letter. Each envelope looks the same. Everyone in both Room A and Room B will randomly pick up an envelope. The person in Room A who chooses the tag with the same letter as yours will be your Actor 1.

Items on your table: 10 E\$ bills (Your endowment).

Items on Actor 1's table: 10 E\$ bills (Actor 1's endowment), two decision sheets (one for Actor 1 and one for Actor 2, as shown below).

Sample Decision Sheets

<p>Decision Sheet Copy for Actor 1</p> <p><u>Actor 1:</u> 1. I decide to transfer _____ E\$ to Actor 2 2. My desired back-transfer amount: _____ E\$</p> <p><u>Actor 2:</u> I decide to transfer _____ E\$ back to Actor 1</p>
--

<p>Decision Sheet Copy for Actor 2</p> <p><u>Actor 1:</u> 1. I decide to transfer _____ E\$ to Actor 2 2. My desired back-transfer amount: _____ E\$</p> <p><u>Actor 2:</u> I decide to transfer _____ E\$ back to Actor 1</p>
--

Actor 1 will make his/her decision at his/her seat by filling in both of the decision sheets. Actor 1 will leave the number of E\$ bills he/she wants to transfer in the envelope he/she picked up, but keep the tag, which will help the experimenter to return the envelope to him/her later. When Actor 1 is done, the experimenter will go to his/her seat, check whether all necessary information is on the decision sheets, then triple the transferred E\$ bills. The experimenter will also record the letter of his/her tag on the back of the decision sheets so that his/her envelope can be given to his/her Actor 2 who has the tag with the same letter. The experimenter will then put the decision sheets and tripled number of E\$ bills in his/her envelope and collect it. After every Actor 1 has finished, the experimenter will take all the envelopes to Room B.

In Room B, the experimenter will first randomly choose half of the envelopes. For each envelope, the experimenter will take out both Decision sheets, and put a Payoff-Cut sticker (as shown below) on both decision sheets and put it back into the envelope. Similarly, for the other half of the envelopes, the experimenter will put No Payoff-Cut sticker on both decision sheets.

Payoff-Cut sticker

Randomly Determined Payoff-Cut
 If Actor 2's back-transfer is less than Actor 1's desired back-transfer amount, there will be a conditional payoff cut of 4 E\$ on Actor 2:
Yes

No Payoff-Cut sticker

Randomly Determined Payoff-Cut
 If Actor 2's back-transfer is less than Actor 1's desired back-transfer amount, there will be a conditional payoff cut of 4 E\$ on Actor 2:
No

After the conditional payoff-cut has been randomly assigned, the experimenter will give each Actor 2 his/her Actor 1's envelope. When you get the envelope, the sticker on the decision sheets will tell you whether you have been randomly assigned to the conditional payoff-cut. Based on this and Actor 1's decision, you will decide how much to transfer back to Actor 1. You will write the amount you want to transfer on both decision sheets, and also place the E\$ bills you want to transfer to Actor 1 in the envelope.

Raise your hand when you are done. The experimenter will go to your seat, check whether all the necessary information is on the decision sheets and then put Actor 1's copy of the decision sheet in the envelope and collect it. You will keep your copy of the decision sheet. Don't show anybody else your decision sheet.

After all Actor 2s have finished, the experimenter will take all the envelopes to Room A. The experimenter will explain the Conditional Payoff Cut to the Actor 1s at this time. *This is the first time that Actor 1s will learn about the Payoff Cut.* The experimenter will return to Actor 1 his/her envelope with the decision sheet and the back-transfer E\$ bills from Actor 2 inside. Each Actor 1 will be called, one by one, to the experimenter. The experimenter will calculate his/her final earnings and pay him/her privately. Then Actor 1 will exit the lab. After all Actor 1s have left, the experimenter will call Actor 2s one by one. When called, you will go to the experimenter with your decision sheet. The experimenter will calculate your earnings and pay you privately.

Throughout this experiment, you won't meet any Actor 1 in room A.

End of Instructions

Please raise your hand to indicate that you are finished reading these instructions. When you're done, an experimenter will give you a short quiz to ensure that you understand how you make decisions.