ThE Papers 06/13

Departamento de Teoría e Historia Económica

Universidad de Granada

An investment game with third-party intervention

Gary Charness (UCSB), Ramón Cobo-Reyes (UGR), Natalia Jiménez (UGR)

An investment game with third-party intervention*

Gary Charness
University of California Santa Barbara

Ramón Cobo-Reyes
University of Granada

Natalia Jiménez
University of Granada and University of Alicante

November 18, 2006

Abstract: This paper explores the effect of the possibility of third-party intervention on behavior in a variant of the Berg, Dickhaut, and McCabe (1995) "Investment Game". A third-party's material payoff is not affected by the decisions made by the other participants, but this person may choose to punish a responder who has been overly selfish. The concern over this possibility may serve to discipline potentially-selfish responders. We also explore a treatment in which the third party may also choose to reward a sender who has received a low net payoff as a result of the responder's action. We find a strong and significant effect of third-party punishment, in both punishment regimes, as the amount sent by the first mover is more than 60% higher when there is the possibility of third-party punishment. We also find that responders return a higher proportion of the amount sent to them when there is the possibility of punishment, with this proportion slightly higher when reward is not feasible. Finally, third parties punish less when reward is feasible, but nevertheless spend more on the combination of reward and punishment when these are both permitted than on punishment when this is the only choice for redressing material outcomes.

Keywords: Trust, punishment, third-party intervention, responsibility-alleviation

JEL Classifications: A13, B49, C91, D63

*We acknowledge comments from Pablo Brañas, Maria Paz Espinosa, Francisco Lagos and Ana Moro. Natalia Jimenez and Ramon Cobo-Reyes are grateful to CEA (SOCH2.05/43) and Generalitat Valenciana (GV06/275) for financial support.

1. Introduction

One must often take a risk in economic environments in order to secure possible gains. Purchasers order goods and presume that the quality when delivered will be suitable; principals hire agents and may not be able to effectively regulate the agent's behavior. The importance of *trust* among individuals has been the topic of considerable discussion in the recent literature. Arrow (1974) pointed out that trust must be required in much economic activity in order to have mutual gains in exchanges. La Porta, López-de-Silanes, Shleifer and Vishny (1997) show that trust has a positive effect on the government effectiveness and also find that trust is associated with lower inflation and with higher per capita GNP growth. Other authors such as Coleman (1990), Fukuyama (1995) or Gambetta (1998) find that trust has a high effect on the functioning of the institutions of a society.[1]

Berg, Dickhaut and McCabe (1995) introduce the Investment Game, in which a first mover can pass some or all of his or her endowment to a responder, who receives three times the amount sent by the first mover. The responder then selects an amount to pass back to the first mover. The behavior of the first mover and the responder can be seen as proxies for 'trusting' and 'trustworthy' behavior, respectively.[2] This game has been used by many authors in the literature trying to obtain the level of trust in the populations analyzed.[3] In general, the amount

[1] Others authors point out the importance of trust in e-commerce. Hoffman, Novak and Peralta (1999) show that almost 95% of consumers have declined to provide personal information to web sites and 63% of them argued that they did not trust those collecting the data. They also show that consumers do not trust Web providers to engage in exchanges, involving money, with them. Mutz (2005) examines the impact of social trust on participation in e-commerce.

[2] These proxies are imperfect, as people may have other motivations for their choices. For example, a first mover may have a desire to increase the total payoff of the group (see Charness and Rabin 2002); similarly, a responder may dislike inequality (rather than being concerned with being trustworthy *per se*) and so may return some to the first mover for this reason. Nevertheless, in this paper we consider a first mover who sends a positive amount to be trusting and a responder who returns at least as much as what was sent to be trustworthy.

[3] See for example Barr (2003), Cox (2004) or Gneezy, Güth and Verboven (2000), among others.

passed by the first mover is less than optimal, so that finding a method to increase the level of trust of the population could open new possibilities of improving economic transactions and would corresponding benefits to society.

The question is then, how can we increase trust experimentally? Berg *et al.* (1995) find that subjects behave differently in a one-shot game if they know the results of earlier games; these results indicate that expectations have an effect on behavior. Barr (2003) shows that trusting behavior is motivated by expectation of trustworthiness. Her analysis shows that altruistic motivations and the desire of "community-build" also appear as possible explanations of trusting behavior. Cox (2002), comparing results in an investment game and in a dictator game, gives evidence that there is significant expectation of trustworthiness in the investment game.

This paper considers a simple modification of the Investment Game, in which a third-party, whose material payoff is unaffected by choices by the first mover and the responder, can choose to sacrifice some of his or her material payoff to punish a responder who presumably returns an insufficient amount to the first mover. The aim of our study is to explore whether introducing the presence of a third-party punisher who will assess other players' performance increases trusting behavior of individuals, based on the above idea that expectations of trustworthiness play an important role in trust.

The punishment performed by uninvolved third parties is not trivial or incidental. Why do people punish others' behavior at a cost for themselves? The economics literature has developed some explanations for this kind of behavior. Levine's (1998) model rests on the assumption that there is a willingness to punish individuals with spiteful preferences. Fehr and Fischbacher (2004) show that third-party sanctions are driven by negative emotions and negative fairness judgments towards norms violators. Gintis, Bowles, Boyd and Fehr (2003) and Fehr, Fischbacher and Gächter (2002) give empirical evidence of what they call "strong reciprocity", the predisposition to cooperate with others and to punish those who violate the norms of cooperation at a personal cost. There are also other possible explanations for this behavior, such as fairness or equality (see Fehr and Schmidt 1999, Charness and Rabin 2002 or Rabin 1993).

To the best of our knowledge, no paper has analyzed the effect of third-party punishment on trusting behavior. We introduce two variants, one in which the third party can only choose whether or not to punish the responder or not, and one in which the third party can choose to either punish the responder, reward the first mover, or do neither (or both). Reward in this framework has been previously considered only in relation to second-party punishment (players who make the punishment decision are involved in the game).[4]

Our results in both variants show that the existence of a person who can observe and punish the violation of the distribution norm (see Fehr and Fischbacher 2004) significantly and substantially increases the amount of money sent by first movers in our population.[5]

We find little difference in the amount sent by first movers when reward is feasible (compared to when only punishment is feasible). Responders return a significant higher amount when third parties can intervene than in the control treatment. The amount sent back is somewhat lower when reward is feasible; perhaps this stems from responders perceiving that some third parties might wish to increase the total social payoffs and thus reward impoverished

---

[4] See for instance Sefton, Shupp and Walker (2002), Fehr & Gächter (2003), Sutter, Haigner and Kocher (2006), Eckel & Grossman (1996), and Ottone (2005).

[5] The Fehr and Fischbacher (2004) study has some similarities to ours, but the game analyzed is different and they did not use reward to enhance cooperation. In one of their treatments they run a Prisoner's dilemma with third-party punishment in two stages. In the first stage, players A and B, who were endowed with 10 points, had only two choices: keep or transfer all points. In the second stage, an external observer endowed with 40 points has the opportunity to punish A and/or B with a ratio of 3 to 1. They focus their analysis in punishers' behavior and found that almost half of C players punished the defector when the other player has cooperated.

first movers rather than punish misbehaving responders. There is also evidence that the level of punishment, in general, increases with increasing difference between first-mover and responder profits. Finally, results show that, although the third party sacrifices more in total when both reward and punishment are feasible, the level of punishment is higher when only punishment can be chosen.

The remainder of the paper is organized as follows. Section 2 briefly describes some of the relevant literature and presents our hypotheses. We present the experimental design and procedures in section 3. The main results are reported and we offer some discussion in section 4. We conclude in section 5.

2. Previous literature & hypotheses

The Berg *et al.* (1995) Investment Game features two types of players, A and B. Both A and B receive an endowment of $10. In the first stage, A can pass some, all, or none of his or her endowment to the paired B. Each dollar sent to Player B is tripled. In the second stage, after observing the amount transferred by A, B decides how many dollars to send back to A, keeping the remainder. The amount B sends back to A is not tripled. If we presume selfish preferences, the subgame-perfect equilibrium prediction for this game is for A to send nothing and for B to return nothing. It is clear that this outcome is not Pareto-efficient, and that the only Pareto efficient allocations occur if and only if A passes his or her entire endowment to B.

Berg *et al.* (1995) found that on average A players sent $5.16 and B players returned $4.66. Thus, on average, A's lose money by sending positive amounts. While B's who receive more do tend to pass more, the proportion returned by B's is uncorrelated with the amount B's receive in their No History treatment; however, there is a positive correlation when Social

History has been provided to the participants, who then have more well-developed expectations about behavior.[6] Expectations clearly matter; since the presence of a third-party punisher can affect these expectations, it seems reasonable to expect this presence to affect behavior.

The Investment Game has been extensively analyzed in the experimental literature in a search for the motivations behind trust.[7] However, the possibility of third-party punishment in this game has not yet been studied. Although many papers have investigated second-party punishment in experimental games,[8] only a few papers analyze third-party sanctions in response to unfair behavior, supporting the idea that those kinds of players do punish in a wide variety of games and contexts, despite the fact that third-party punishers must sacrifice their endowment to penalize an action in which they are not involved. Fehr and Fischbacher (2004) study third-party sanctions in a Dictator Game and in a Prisoners' Dilemma. They find that, in the Dictator Game, most third parties punished dictators who transferred less than half of the pie. In the Prisoners' Dilemma, results show that 46% of third parties punish defectors if their partners cooperated.

Based on results obtained by Fehr and Gächter (2000) or Fehr and Gächter (2002),[9] we test the following hypotheses concerning the effect of introducing a third-party punisher in the BDM investment game:

H1: Introducing a third party who can intervene will lead to A's sending more to B's.

H2: B's will return more to A's if there is a third party who can intervene.

---

[6] The Spearman correlation coefficient is $r_s = 0.01$ in the No History treatment, compared to $r_s = 0.34$ in the Social History treatment.

[7] See for instance Croson & Buchan (1999), Cox (2002), Cox (2004), Gneezy *et al.* (2000).

[8] See, for example, Camerer (2003), Güth, Schmittberger and Schwarze (1982), Fehr, Gächter and Kirchsteiger (1997) or Ostrom, Walker and Gardner (1992) among others.

[9] They find that the threat of punishment has an effect on contributions in public-goods games.

For establishing H1, we presume that one of the main considerations people have about sending money in the Investment Game is that they would like to receive back at least the same amount that they have sent, being no worse off. Cox (2002) and Barr (2003) provide results supporting this assumption. The idea behind third-party intervention is that uninvolved players can punish unfair behavior or perhaps also help victims of violations to this 'norm'.

Reward has been principally analyzed in combination with second-party punishment. Sefton *et al.* (2002) compare second-party reward to cooperators versus punishment to defectors in a public goods game and found that reward is less efficient, especially in relation with the decay in time.[10] Andreoni, Harbaugh and Vesterlund (2003) analyze an ultimatum game with an additional stage where the responder can reward or punish the proposer (with some cost). They run four treatments: the baseline, only punishment, only reward, and both punishment and reward. They find that reward and punishment are complements, so the last treatment is the most efficient for enhancing cooperation. Charness and Levine (forthcoming) incorporate both second-party punishment and reward in a design where the final "wage" received by the responder is in part stochastic; even at the same net wage, the responder chooses reward (punishment) more frequently when the first mover has chosen a high (low) wage.

Third-party punishment and reward are combined in Ottone (2005), who proposes a Dictator Game with a second stage where an external observer could punish dictators and/or reward recipients. She shows that the level of punishment is proportional to unfair offers of dictators. But the interaction between the first and second parties is not a strategic environment and there are no efficiency considerations in the Dictator Game.

[10] Fehr & Gächter (2003) also found similar results in a public goods game. Sutter et al. (2006) introduce endogenous choice of the reward or sanction system in a public goods game and they show that cooperation is higher when subjects can choose the system than when it is exogenously imposed.

In order to compare reward and punishment to punishment only, we also run a treatment in which an unrelated player, after observing A's and B's behavior, has the opportunity to reward player A and/or punish player B. Thus, we also test the following hypotheses:

H3: First movers will transfer more when both reward and punishment are feasible than when only punishment is feasible.

H4: Responders will return less when both reward and punishment are feasible than when only punishment is feasible.

To motivate H3, we consider that a first mover may feel safer in transferring a larger amount when there is the additional possibility that the third party can remedy a shortfall from the responder. Regarding H4, a responder may perceive less of a threat, as a third party who cares about efficiency may choose to help the first mover, rather than to punish the responder.[11]

A final operational note is that we decided to have the participants themselves vote for the third parties, rather than choosing them at random. Here the idea was to increase the sense of responsibility for the third parties. The intuition behind this is that when third parties know that they have been *selected* as punishers by the rest of participants in the experiment, they will feel more involved in their task. First, we asked subjects to all stand and show their 'identification numbers'. They then voted for one of these identification numbers. No other information was provided to the subjects before the voting process, as additional personal information could have affected the perceived social distance, which has been shown to have an important effect.[12]

[11] The notion of *responsibility alleviation* (Charness 2000) suggests that responders will make more pro-social choices (return more) when they are fully responsible for the first mover's payoff (in the punishment only case).
[12] See for instance Bohnet & Frey (1999), Charness & Gneezy (forthcoming).

3. Experimental design

We conducted six sessions at UCSB, two for each of the three different treatments, with 16-24 participants per session. Subjects were recruited from a university-wide e-mail database of interested students. All sessions were run in a large classroom divided into two sides and in each row four people sat sufficiently separated. No one was allowed to participate in more than one session. On average, each person received about $13 for a one-hour session.

The first treatment was the standard Investment Game where both players A and B received an endowment of 10 points. The other two treatments featured potential third-party punishment. At the beginning of the second and third treatments, participants were handed out a label with a number printed in it. After explaining the game in detail, people voted for the third parties; those people receiving the most votes became the third parties, with ties broken randomly. The elected third parties then were asked move to the back of the room, maintaining separation among them.

In all treatments, the roles for A and B were randomly assigned with labels drawn from an opaque bag (in the second and third treatments these labels were distributed for the remaining participants after the voting for third parties). Each label had a letter (A, B or C) and a number. Once the roles were determined, A's and B's were seated on opposite sides of the room, with separation between all individuals for privacy.

In treatments 2 and 3, A's and B's were endowed with 10 points, while C's were endowed with 40 points.[13] The conversion rate for points to dollars was 2:1 for A's and B's and 3:1 for C's; each participant knew that the conversion rate for C's could be different than that for

---

[13] We decided to endow C with a sufficiently large amount of points to avoid C comparing his or her points with B amount. Thus, in the most favorable case for B, when A transfers 10 points, C will not punish B in order to decrease the difference in payoffs between B and C (see Fehr & Fischbacher 2004).

A's and B's, but were only told their own conversion rate. C's could choose to punish (or reward, in treatment 3); each point sacrificed by C decreased B's payoff by three points or (with reward) increased A's payoff by three points, and this was common information.

There were three rounds in each session, with no feedback given. Subjects retained their role in all rounds, but they were matched with different partners in each one.[14] Participants were told at the outset that they would be paid for only one of the three rounds, chosen randomly at the end of the session by rolling a die. The experiment was structured in the following way. First, decision sheets were passed out to all participants. A's decision sheets had the label they had drawn written in the top left corner of each decision sheet. A's passed their decision sheets face down toward the center aisle and these were then collected. We then wrote the label for the corresponding B in the top right corner of A's decision sheet, and cut off the corner where the A player's label was written. These decision sheets were distributed to the corresponding B's.

Each B observed what A has chosen and then made a decision on a separate decision sheet with B's label in the top right corner; this was passed face down to the center aisle and collected, along with A's decision sheet. We then wrote the identification of the corresponding C in the top left corner of both decision sheets and cut off the top right corner of the decision. Next, we distributed both sheets to C, who observed A's and B's decisions and made choices about punishment (and reward, in treatment 3). These were collected face down and placed in a folder. In this manner, our procedural design comes close to a "double-blind" protocol. This process was repeated in each of the three rounds. When a round was selected for payment at the end of the session, we opened the corresponding folder and calculated actual payoffs.

[14] This was common information among subjects.

4. Results

In this section, we first present the results for A's, then the responses by B's, and then the choices made by C's. In the process, we also discuss how the data either support or reject our hypotheses. As a summary, we can say that the presence of a third party has a strong effect on both the amounts sent by A's and the amounts returned by B's. C's are more prone to sacrifice points when there is more inequity between the corresponding A and B payoffs at the time of C's decision, and C sacrifices slightly more when both punishment and reward are feasible.

In treatments 1, 2, and 3, we had 16, 14 and 12 groups of subjects respectively. As we run three rounds for each experiment, we have tested if the three different rounds for A, B and C players could be pooled in a single sample (of 48, 42 and 36 observations respectively). We have used a single-factor within-subjects analysis of variance test that is appropriate for two or more dependent populations. On this basis, we pool results across rounds, as we do not find substantial differences over time.[15]

Amounts sent by A's

Table I shows the average amount of points sent by player A in each treatment.

Table 1: A's choices, by treatment

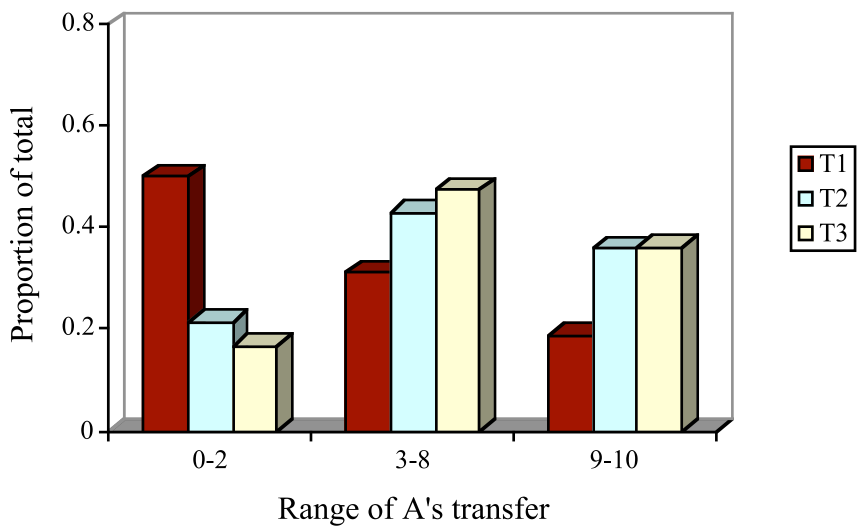
	Treatment 1	Treatment 2	Treatment 3
Mean transfer	3.73	6.02	6.14
Observations	48	42	36

[15] Pooling across rounds seems particularly innocuous for A's, as they have received no feedback at all at the time of their decisions in rounds 2 and 3.

We see that A's transfers are more than 60% higher when there is the possibility of third-party punishment, and that there is very little difference in A's behavior depending on whether the third party can also choose to reward A.

Figure 1 shows the distribution of low, intermediate, and high A transfers in each treatment.

Figure 1 - A's transfers, by treatment



Visually, we see that these results are driven by the fact that low transfers are much more frequent without third-party punishment, and high transfers are considerably more common when third-party punishment is permitted.

A conservative Wilcoxon-Mann-Whitney ranksum test using each individual's average transfer as one observation confirms that the difference between treatment 1 and treatment 2 is statistically significant ($Z = 1.98$, $p = 0.024$, one-tailed test), as is the difference between treatment 1 and treatment 3 ($Z = 2.05$, $p = 0.020$, one-tailed test). A regression with standard errors clustered on the individual confirms the statistical significance ($Z = 1.96$ and $Z = 1.97$, respectively). The observed behavior therefore confirms **H1**, that A's transfer more when there is the possibility of third-party punishment. This is the main result of the paper. On the other hand, the lack of difference between A's transfers in treatments 2 and 3 (the Wilcoxon ranksum test gives $Z = 0.13$, n.s.) goes against **H3**, the hypothesis that A's will transfer more when reward is feasible.

Amounts sent by B's

The effect of potential third-party intervention on responder behavior could be analyzed from several alternative points of view. For instance, while it might seem obvious to simply consider the absolute amounts transferred from B to A, one concern is that we are comparing treatments where, in one of them, the amount sent by A is much lower than in the others. One might therefore expect B's transfers to be higher in treatments 2 and 3, even if B's had no concern with possibly being sanctioned. Another possibility may be to analyze B's behavior in terms of the proportion of the original A transfer. Yet this approach also has problems – when A's transfers are low, then it is inexpensive for B to return a high percentage of the points transferred by A. However, if A's transfers are high, then B must give up a larger sum of money to achieve this same proportion.

As with nearly all studies investigating behavior in a sequential prisoner's dilemma game such as ours, there is a strong positive relationship between the amount sent by the first mover and the responder's choice; in all treatments, the correlation between A's transfer and B's transfer is positive and highly significant.

Table 2 shows the mean transfer by B's in each treatment, as well as the average ratio of B's transfer to the amount sent by A.[16]

[16] Simply calculating the aggregate ratio in each treatment will give excessive weight to observations where a higher amount was passed by A. We drop the observations where A transferred 0, as the ratio is undefined in that case.

Table 2: B's choices, by treatment

	Treatment 1	Treatment 2	Treatment 3
Mean transfer	4.65	8.33	6.99
Mean ratio to A's transfer	1.11	1.40	1.30
Observations	48	42	36

We see that while B's transfer substantially more to A's in treatments 2 and 3, the difference in the average ratio is only a modest one. Nevertheless, the relationship between these average ratios is an intuitive one: Introducing a third party increases B's responsiveness to A's transfer in both treatments 2 and 3, but this increase is smaller in treatment 3 when the third party can choose to reward A as well as to punish B.

For the reasons explained above, we might analyze B's transfers as a variable conditioned on the amount sent by A. Thus, we consider the General Lineal Model:

$$E(Y|X = x_i) = a + bx_i,$$

where $Y|X = x_i$ is the amount B returns to A conditioned to the fact that A has transferred x_i points to B. Once we run the linear regression, we then compute the predicted values of the dependent variable, $\hat{E}(Y|X = x_i)$. Finally, we analyze, using statistical tests, whether the means of predicted values for B's transfers differ across treatments.

If we compare the mean transfers for each individual in treatment 1 with those in treatment 2 or 3, a Wilcoxon ranksum test confirms that B's transfer is higher when there is a third party than when there is not ($Z = 3.726$, $p = 0.000$, T1 vs. T2 and $Z = 2.860$, $p = 0.002$, T1 vs. T3, one-tailed tests). Therefore, we find support for **H2**.

We find some modest support for **H4**. Although the average A transfer is a bit higher when only punishment is feasible, Table 2 indicates that the average transfer in treatment 2 is higher than in treatment 3. A Wilcoxon ranksum test (on the predicted values for B's transfer) finds a marginally-significant difference in the levels of B's transfer across treatments 2 and 3 (Z = 1,339, p = 0.090, one-tailed test). This result is compatible with B's responsibility alleviation, given that in treatment 3 A's profit not only depends on B's transfer but also on C's reward.

Amounts sacrificed by C's

While the behavior of the elected third parties is not the main focus of our study, we nevertheless briefly present these results. Table 3 shows the average amount of money spent by third parties in treatments 2 and 3.

Table 3: Average third-party choices, by treatment

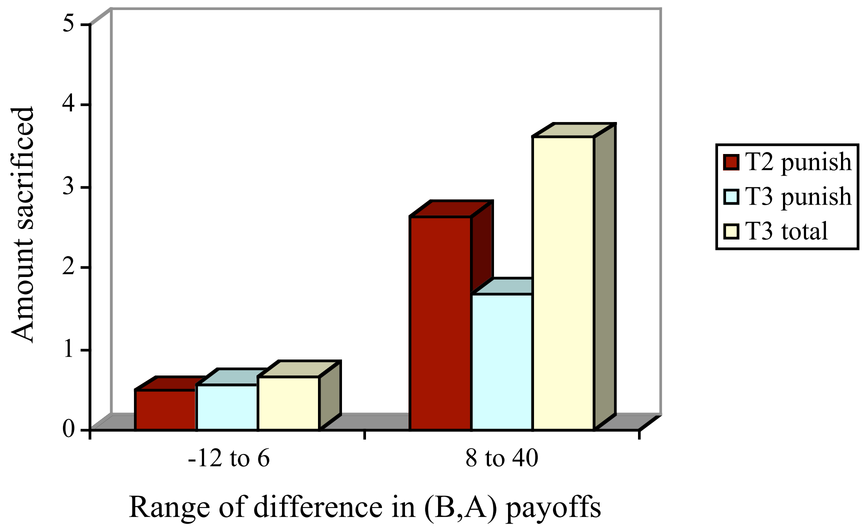
	Treatment 2	Treatment 3
Amount spent to punish	1.45	1.03
Amount spent to reward	-	0.86
Total amount spent	1.45	1.89
Observations	42	36

We observe 41% more points spent on punishment when this is the only tool available. On the other hand, the total expenditure on intervention is 30% higher when both reward and punishment are feasible. Notwithstanding reward is the most efficient action in treatment 3, C prefers to sacrifice 20% more points to punish than to reward. These results seem fairly intuitive. Since the third parties tend to intervene only when B misbehaves, we have relatively few 'useful' observations.[17] Thus, statistical tests have little power here and neither rank-sum tests nor regressions find either difference to be significant.

Nevertheless, there is a clear relationship between third-party behavior and the combination of A and B choices, as shown in Figure 2 below (the two ranges were chosen to give roughly equal numbers of observations):

[17] There are many cases where B returns a considerable proportion of what B receives from A, so that C does not punish. In addition, C never punishes B when A sends 0.

Figure 2 - Average third-party sacrifice



There is little sacrifice by C's when the difference between B's payoff and A's payoff is relatively small, but a sizable amount of punishment when this difference is larger. The amount of reward is a bit larger than the amount of punishment when the difference in payoffs is larger. It is clear that C's are not acting in a random fashion.

Table 4: Regressions on third-party intervention

	Dependent variable		
Independent variable	(1) Punishment	(2) Reward	(3) Total
B payoff – A payoff	0.065*** (.019)	0.087*** (.017)	0.118*** (.018)
Treatment 3	-0.601 (.604)	-	0.116 (.670)
Constant	0.941** (.436)	-0.060 (.344)	0.524 (.476)
N	78	36	78
R^2	.112	.425	.291

Standard errors are clustered by individual and are in parentheses.
***, **, and * indicate significance at $p = 0.01, 0.05$, and 0.10, respectively

Table 4 confirms the visual results in Figure 2, as the difference between B's payoff and A's payoff is highly significant and appears to drive third-party intervention; when this disparity is too large, some third parties sacrifice payoffs.[18] As mentioned above, we find no treatment effect. On average, an increase of 8.5 points in this difference leads to an increase of one unit in total intervention.

5. Conclusion

In this paper, we investigate whether the possibility of intervention by an external third party can affect the behavior of individuals in a form of the Berg *et al.* (1995) Investment Game. In particular, we are interested in whether the amount passed by the first mover, which is the source of social gains in this game, is increased when a third party can make a sacrifice to bring payoffs closer to balance. We consider one variation in which punishment is the only feasible intervention and another variation in which both reward and punishment are possible.

We find that first movers are willing to 'trust' more when a third party is present, presumably through the belief that the responder is less likely to be selfish. The amount passed increases by over 60% over the no-punishment baseline in both treatments with feasible third-party intervention.

Regarding the amount returned by responders, we find that third party intervention also affects positively this amount. Nevertheless, the amount sent back is higher when only

[18] Note that it is unnecessary to run a regression in which A's payoff is an independent variable, given that the difference of payoffs between A's and B's is very high if and only if A's send many points and B's return a very low amount. Thus, in this setting we cannot distinguish whether C's care about inequality among A's and B's payoff or if they care about unfair B's behavior.

punishment is feasible than when both reward and punish are possible. The threat of punishment might be perceived weaker by responders when reward is also a possible action.

The third parties are in fact willing to spend some of their payoff to redress payoff disparities, as the amount spent is strongly correlated with the amount by which the responder's payoff exceeds the first-mover's payoff. There is more punishment when this is the only available sacrifice, but the total sacrifice is greater when reward is also an option. While responders return more in the treatments with a third party, this amount is lower when reward is also feasible. As we have controlled for the effect of the amount received by responders, we may rule out the idea that this higher return is due to the fact that first movers have passed more in the first place.

Our study demonstrates the relevance of external third parties in a sequential game. It seems that some people in the experimental population expect that an outsider will be willing to incur some personal cost to punish the greedy or to help trusting victims. In our environment, the result is more trusting behavior. To the extent that the laboratory is indicative of behavior in the field,[19] perhaps there is scope for policy mechanisms such as third-party intervention to improve trust more generally.

---

[19] Since emotion in the field tends to run higher than emotion in the laboratory, one might even consider laboratory punishment understates punishment in the field.

References

Andreoni, J., Harbaugh, W. and L. Vesterlund (2003) "The carrot or the stick: Rewards, punishments and cooperation". *American Economic Review* 93: 893-902.

Arrow, K. (1974) *The Limits of Organization.* NY: Norton.

Barr, A. (2003) "Trust and Expected Trustworthiness: Experimental Evidence from Zimbabwean Villages". *The Economic Journal*, 113: 614-630.

Berg, J., Dickhaut, J. and K. McCabe (1995) "Trust, Reciprocity and Social History". *Games and Economic Behavior,* 10: 122-142.

Bonhet, I. and B. Frey (1999) "Social Distance and Other Regarding Behavior in Dictator Games: Comment" *American Economic Review*, 89: 335-339.

Camerer, C. (2003). Behavioral game theory. Princeton, NJ: Princeton University Press.

Charness, G. (2000) "Responsibility and Effort in an Experimental Labor Market". *Journal of Economic Behavior and Organization*, 42: 375-384

Charness, G. and U. Gneezy (forthcoming) "What's in a Name? Anonymity and Social Distance in Dictator and Ultimatum Games". *Journal of Economic Behavior and Organization.*

Charness, G. and D. Levine (forthcoming) "Intentions and Stochastic Outcomes: an experimental study". *Economic Journal.*

Charness, G. and M. Rabin (2002) "Understanding social preferences with simple tests". *The Quarterly Journal of Economics* 117: 817-869.

Cox, J.C. (2004) "How to identify trust and reciprocity". *Games and Economic Behavior* 46: 260-281.

Cox, J.C. (2002) "Trust, Reciprocity, and Other-Regarding Preferences: Groups vs. Individuals and Males vs. Females". Chapter 14 in Rami Zwick and Amnon Rapoport, (eds.), Experimental Business Research, Kluwer Academic Publishers.

Coleman, J. (1990) "Foundations of Social Theory". Cambridge, MA: Harvard University Press.

Croson, R. and N. Buchan (1999) "Gender and Culture: International Experimental Evidence from Trust Games". *American Economic Review, Papers and Proceedings,* 89: 386-391.

Eckel, C. and P. Grossman (1996) "The relative price of fairness: gender differences in a punishment game" *Journal of Economic Behavior and Organization*, 30: 143-158.

Fehr, E. and S. Gächter (2002) "Altruistic Punishment in Humans". *Nature*, 415: 137-140.

Fehr, E., Gächter, S. (2003), "Self-governing solutions to collective action problems: Rewards versus punishments". Preliminary draft.

Fehr, E. and U. Fischbacher (2004) "Third party punishment and social norms". *Evolution and Human Behavior*, 25: 63-87.

Fehr, E., Gächter, S. and G. Kirchsteiger (1997) "Reciprocity as a contract enforcement device". *Econometrica*, 65: 833–860.

Fehr, E. and K. Schmidt (1999) "A theory of fairness, Competition and Cooperation". *The Quarterly Journal of Economics* 114: 817-868.

Fehr, E. and S. Gächter (2000) "Cooperation and Punishment in Public Goods Experiments". *American Economic Review*, 90: 980-994.

Fehr, E., U. Fischbacher and S. Gächter (2002) "Strong Reciprocity, Human Cooperation and the enforcement of social norms". *Human nature*, 13: 1-25.

Fukuyama, F. (1995) “Trust”. NY: Free Press.

Gambetta, D. (1988) “Trust: Making and Breaking Cooperative Relations”. Cambridge: Blackwell.

Gintis, H., R. Bowles, R. Boyd and E. Fehr, (2003) “Explaining altruistic behavior in humans”. *Evolution and Human Behavior* 24: 153-172.

Gneezy, U., W. Güth, and F. Verboven (2000) “Presents or Investments? An Experimental Analysis”. *Journal of Economic Psychology*, 21: 481-493.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). “An experimental analysis of ultimatum bargaining”. *Journal of Economic Behavior and Organization*, 3: 367–388.

Hoffman, D., T. Novak and M. Peralta (1999) “Building consumers trust Online” *Communications of the ACM*, 42: 80-85.

La Porta, R., F. López-de-Silanes, A. Shleifer and R. Vishny. (1997) “Trust in Large organizations” *American Economic Review: Papers and Proceedings*, 87: 333-338.

Levine, D. (1998) “Modeling altruism and spitefulness in experiments” *Review of Economic Dynamics*, 1: 593-622.

Mutz, D. (2005) “Social Trust and E-Commerce: Experimental Evidence for the effects of Social Trust on individual Economic Behavior”. *Public Opinion Quarterly*, 69: 393-416.

Ostrom, E., Walker, J. and R. Gardner (1992). “Covenants with and without a sword: Self-governance is possible”. *American Political Science Review*, 86: 404–417.

Ottone, S. (2005) “Transfers and altruistic punishment in Solomon's Game experiments”. Mimeo. Department of Public Policy and Public Choice – POLIS. Working paper No 57.

Rabin, M. (1993) “Incorporating Fairness into Game Theory and Economics”. *American Economic Review*, 83:1281-1302.

Sefton, M., Shupp, R. and J. Walker (2002) “The effect of reward and sanctions in the provision of public goods” University of Nottingham, Working Paper.

Sutter M., S. Haigner and M. Kocher (2006) “Choosing the Stick or the Carrot? Endogenous Institutional Choice in Social Dilemma Situations”. CEPR Discussion Paper No. 5497.