Economics Letters 116 (2012) 429–431

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Fairness, spite, and intentions: Testing different motives behind punishment in a prisoners' dilemma game☆

Charlotte Klempt*

Eberhard Karls University of Tübingen, Germany
Institute for Applied Economic Research, Germany

ARTICLE INFO

Article history:
Received 15 February 2012
Received in revised form
5 April 2012
Accepted 20 April 2012
Available online 27 April 2012

JEL classification:
A13
D63
C92

Keywords:
Intentions
Fairness
Spitefulness
Sanctioning
Cooperation

ABSTRACT

This paper differentiates between three motives behind punishment in a social dilemma: minimizing inequalities, retaliation against unfair acts, and spitefulness. The experiment shows that cooperators and defectors differently respond to intentions and thereby substantiates Falk et al. (2005)'s findings.

1. Introduction

Recent theoretical studies incorporate the fact that not only material self interest drives people's behavior (Camerer, 2003). While Bolton and Ockenfels (2000) and Fehr and Schmidt (2000) assume that the desire to adjust payoff differences among players defines people's fairness concerns, Dufwenberg and Kirchsteiger (2004) and Rabin (1993) focus on the fact that the actor's intentions might play a crucial role when people judge the fairness of an action. Experiments confirm that both — the outcome and intention based fairness notion — behaviorally matter (Offerman, 2002; Falk et al., 2003; Charness and Levine, 2007; Falk et al., 2008). While there are studies that show that different players incorporate different fairness notions (Falk et al., 2005; Anderson and Putterman, 2006; Sánchez-Pagés and Vorsatz, 2007), there is no study as yet testing whether distinct player types differently account for other players' intentions. The present study tests and extents the evidence given by Falk et al. (2005)'s study (henceforth FFF) in order to investigate these potential different types.

The experimental design relies on FFF's three players' prisoners' dilemma (PD) and additionally alters the intentionality of the player's decisions. FFF's experiment analyzes why subjects punish others in a PD. In the first stage, players can decide whether to cooperate or to defect. The second stage allows them to punish both partners given the others' PD decision. The punishment costs alter in two treatments: In the high sanction treatment, punishing allows players to improve their relative payoff standing as opposed to the low sanction treatment.

While cooperators only punish other defectors, defectors punish both — cooperators and defectors — to the same extent. Cooperators equally punish defectors in the low and high sanction treatment and do not punish other cooperators at all. Defectors only punish others in the high sanction treatment and do not punish at all in the low sanction treatment.

FFF conclude that fairness motives drive cooperators' sanctioning and that defectors' sanctioning is driven by spite. Cooperation and punishment of defectors is consistent with fairness

☆ I am grateful to the Max Planck Society for financial support through the IMPRS Uncertainty. I thank the anonymous referee(s), Werner Güth, Oliver Kirchkamp, Kerstin Pull, and (former) colleagues for helpful comments. A large part of this paper has been written at the Max Planck Institute of Economics in Jena.

* Correspondence to: Eberhard Karls University of Tübingen, HRM and Organization, Nauklerstr. 47, 72074 Tübingen, Germany. Tel.: +49 7071 29 78196; fax: +49 7071 29 5077.

E-mail addresses: charlotte.klempt@uni-tuebingen.de, charlotte.klempt@iaw.edu.

0165-1765/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2012.04.048

approaches. FFF here discuss two fairness principles that might explain cooperators' sanctioning: fairness-driven sanctioning that is motivated by the desire to retaliate against unfair acts and fairness-driven sanctioning that aims at minimizing payoff inequalities. As cooperators impose the same payoff reductions on defectors even if they cannot reduce payoff inequalities, FFF conclude that retaliation drives their sanctioning. In turn, defectors only punish in the high sanction treatment if it improves their relative payoff standing. As they equally punish cooperators and defectors, inequity aversion can not explain their behavior and punishment seems to be driven by spite.

The present study is designed to test the conclusions drawn by the findings offered by FFF in two ways. *First*, the paper tests whether spitefulness solely explains the defectors' sanctioning and, *second*, whether the desire to retaliate exclusively explains the cooperators' punishment. I test these two explanations by considering the fact that the desire to retaliate against unfair acts involves the attribution of intentions. That is, if cooperators sanction in order to retaliate against deliberate defection, they should not punish unintentional defection at all. Cooperators solely concerned about distributional inequalities, however, should ignore intentions and only focus on outcomes. In turn, spiteful punishment by defectors should disregard others' intentions as it solely aims at increasing payoff differences.

In the present paper, I alter the intentionality of decisions by replacing some PD decisions with a random draw. The results reveal that only cooperators but not defectors decrease their punishment in response to unknown intentions. Cooperators punish defectors much less if the latter can not be hold responsible for that decision. While retaliation partly seems to drive cooperators sanctioning, the large amount of punishment directed towards unintended defection supports additional inequity concerns. Defectors in turn do not respond to intentions and results further confirm that their sanctioning might purely arise from spiteful motives. In addition, the results reveal a surprisingly low fraction of subjects that take their partner's intention into account.

2. Experimental setting

The experiment employs a three-player PD with punishment opportunity relying on the parameter values of FFF but replacing the decision of one PD player with a random decision.

In the first stage of the experiment, subjects decide simultaneously whether to cooperate or to defect. The payoff consequences of the PD decisions are depicted in Table 1.

After all PD decisions are made, the decision of one randomly chosen subject is replaced by a chance move. The PD players do not know in advance whose decision will be replaced. Nature will choose to cooperate or to defect with equal probability.

In the second stage, one randomly chosen PD player can punish the other two by assigning them deduction points. The punishing player is never the subject with the random PD decision. Hence, the punishing player always faces both types of PD players — one player whose decision was given by chance and one whose decision was made by choice.

The punishing player can assign up to 15 deduction points to each of his PD partners. Each deduction point reduces his own payoff by one point and the punished players payoff by 2.5 points. Thus, punishment is costly for the punishing and the punished subject.

The experiment uses the strategy method: players make their choices for all roles continent on each possible partner's choice. They do not know whether they are the actual punishing player or whether their PD decision has been switched. All subjects assign deduction points before learning whether they are the punishing player and before knowing the PD partners' first stage choices.

Table 1
Payoff to Player i.

	Both other players defect	One of the the other players cooperates	Both other players cooperate
Player i defects	20	32	44
Player i cooperates	12	24	36

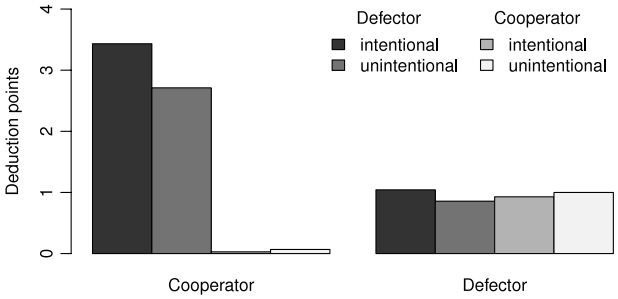

Fig. 1. Average deduction points assigned.

Hence, subjects assign deduction points for all possible choice combinations made by both partners. These include four possible cases: both of the other partners defected, both cooperated, the second player cooperated while the third defected, and vice versa. Subjects know which of both players will represent the one whose decision has been switched. After all decisions are made, the roles are assigned and payoffs calculated: one subject's PD decision will be switched and punishment points of one subject will be deducted.

The experiment was conducted in the laboratory of the Max-Planck-Institute of Economics in Jena, Germany. Subjects were students of the University of Jena, and were recruited from an online subject pool. Subjects were anonymously paired, and their identities were never revealed to one another. Each subject only played one round of the game. I conducted three sessions for a total of 87 subjects. I used similar instructions as FFF with small modifications according to changes in the experimental design. Subjects filled out a questionnaire testing comprehension of the rules. Each point earned in the experiment was exchanged for €0.2. Subjects earned on average €9.73 including a participation fee of €4. At the end, subjects received feedback on their roles, others decisions, and payoffs. The experiment was conducted with z-Tree (Fischbacher, 2007).

3. Results

In the PD, sixty percent of the subjects cooperated, while the others defected. Fig. 1 depicts the average deduction points assigned by cooperators and defectors.

Firstly, we observe that defectors equally punish cooperators and defectors irrespective whether the punished subject is responsible for his decision or not. On average, they assign 0.93 points to cooperators and 1.04 points to defectors who deliberately chose their decision and 1.00 points to cooperators and 0.86 points on defectors whose decision was randomly chosen. These choices do not reveal any differences (Wilcoxon signed ranked test, $p > 0.62$; t-test, $p > 0.71$).

These results confirm that defectors punishment might be driven by spite and are in line with the results of FFF. Defectors do not respond to the attribution of intentions and merely aim at increasing their payoff standing.

Cooperators show a different sanctioning pattern. Punishment directed towards other cooperators is almost negligible and does not respond to intentions (Intentional vs. unintentional

Table 2
Percentage of cooperators and defectors who punish.

	Defector[a]		Cooperator	
	Intentional	Non-intentional	Intentional	Non-intentional
Sanctioned subject is a defector	0.157	0.129	0.462	0.365
Sanctioned subject is a cooperator	0.129	0.157	0.01	0.029

[a] The symmetry of frequencies is fortuitous.

cooperation: 0.03 points vs. 0.07 points, Wilcoxon signed ranked test, $p > 0.41$; t-test, $p > 0.42$). In turn, cooperators do punish other defectors and assign significantly more punishment points to subjects defecting by choice than by chance (Intentional vs. unintentional defection: 3.41 points vs. 2.71 points, Wilcoxon signed ranked test, $p < 0.01$; t-test, $p < 0.01$).

These results partly confirm the findings by FFF as cooperators only assign deduction points to other defectors. But do cooperators punish to retaliate against free riding or to decrease payoff inequalities? The results suggest that cooperator's punishment is at least partly driven by the desire to retaliate as they especially want to harm those who deliberately free ride. But, we still observe a substantial amount of punishment directed towards unintended defection as punishment only decreases by 20%. There are two possible explanations: Firstly, cooperators might aim at punishing actual defectors whose decision was switched. Secondly, they might punish in order to decrease payoff inequalities. The first explanation would only hold if cooperators similarly punished unintentional defection and unintentional cooperation. In both cases, they might be confronted with an actual defector. As cooperators only punish unintentional defectors, they presumably aim at decreasing payoff inequalities.

When looking at the percentage of subjects who actually punish in Table 2, the findings reflect the ones of Fig. 1. However, we only observe a 10% decrease when comparing the fractions of cooperators punishing intentional and unintentional defection (46% vs. 36%, Fisher's exact test, $p < 0.1$). This is surprisingly low regarding former studies on intentions (Offerman (2002) finds a decrease of 67%, Charness and Levine (2007) find 39%). These difference may arise due to the distinct experimental setting: Subjects in this experiment play a simultaneous dilemma game where all face the same choice. Former experiments employed a sequential bargaining situation with unequal roles. The former situation might be perceived as procedurally fairer and therefore alter behavior (Bolton et al., 2005).

4. Conclusion

The paper adds to the literature in three ways. First, the experiment shows that there are different subject types that differently correspond to people's intentions. Second, it shows that it is possible to test assumptions regarding people's fairness motives by varying the intentionality of decisions. And third, the experiment suggests that a subject's focus on intentions varies with game type.

References

Anderson, Christopher, Putterman, Louis, 2006. Do non-strategic sanctions obey the law of demand? Games and Economic Behavior 54 (1), 1–24.
Bolton, Gary, Brandts, Jordi, Ockenfels, Axel, 2005. Fair procedures: evidence from games involving lotteries. The Economic Journal 115 (506).
Bolton, Gary, Ockenfels, Axel, 2000. ERC: a theory of equity, reciprocity, and competition. American Economic Review 90 (1), 166–193.
Camerer, Colin, 2003. Behavioral Game Theory. Princeton University Press, Princeton.
Charness, Gary, Levine, David, 2007. Intention and stochastic outcomes: an experimental study. The Economic Journal 117 (522), 1051–1072.
Dufwenberg, Martin, Kirchsteiger, Georg, 2004. A theory of sequential reciprocity. Games and Economic Behavior 47 (2), 268–298.
Falk, Armin, Fehr, Ernst, Fischbacher, Urs, 2003. On the nature of fair behavior. Economic Inquiry 41 (1), 20–26.
Falk, Armin, Fehr, Ernst, Fischbacher, Urs, 2005. Driving forces behind informal sanctions. Econometrica 73 (6), 2017–2030.
Falk, Armin, Fehr, Ernst, Fischbacher, Urs, 2008. Testing theories of fairness–Intentions matter. Games and Economic Behavior 62 (1), 287–303.
Fehr, Ernst, Schmidt, Klaus, 2000. Fairness, incentives, and contractual choices. European Economic Review 44 (4–6), 1057–1068.
Fischbacher, Urs, 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10 (2), 171–178.
Offerman, Theo, 2002. Hurting hurts more than helping helps. European Economic Review 46 (8), 1423–1437.
Rabin, Matthew, 1993. Incorporating fairness into game theory and economics. The American Economic Review 38 (5), 1281–1302.
Sánchez-Pagés, Santiago, Vorsatz, Marc, 2007. An experimental study of truth-telling in a sender–receiver game. Games and Economic Behavior 61 (1), 86–112.