Journal of Economic Behavior & Organization 84 (2012) 753–766

Contents lists available at SciVerse ScienceDirect

Journal of Economic Behavior & Organization

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

An exploration of third and second party punishment in ten simple games☆

Andreas Leibbrandt[a,*], Raúl López-Pérez[b]

[a] Economics Department, Monash University, Clayton 3800, Victoria, Australia
[b] Department of Economic Theory, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 9 March 2010
Received in revised form
17 September 2012
Accepted 19 September 2012
Available online 29 September 2012

JEL classification:
C70
C91
D63
D74
Z13

Keywords:
Fairness
Inequity aversion
Norms
Punishment
Reciprocity
Third parties

ABSTRACT

This paper explores the motivations behind punishment from unaffected third parties and affected second parties using a within-subjects design in ten simple games. We apply a classification analysis and find that a parsimonious model assuming that subjects are either inequity averse or selfish best explains the pattern of punishment from both third and second parties. Despite their unaffected position, we find that many third parties do not punish in an impartial or normative manner.

1. Introduction

Third parties play a crucial role in many institutions: They serve in courts, as referees or arbitrators.[1] The US legal system, for instance, relies on the judgment by juries when it comes to the application of sanctions. Third parties are also important with regard to informal sanctions (Homans, 1961) and, in fact, their interventions seem to be essential in the explanation of norm enforcement, as they are often more numerous than affected second parties (Bendor and Swistak, 2001) or the only parties present and hence their sanctions are potentially more damaging than those from second parties.

☆ The authors greatly acknowledge financial support by the EU Research Network ENABLE. We also want to thank Gary Charness, Martin Dufwenberg, Armin Falk, Ernst Fehr, James H. Fowler, David M. Grether, Daniel Houser, Antonio Martín-Arroyo, Ernesto Reuben, Tatsuyoshi Saijo, Frans van Winden, one associate editor, two referees, and numerous participants at several conferences who provided valuable feedback.

* Corresponding author. Tel.: +61 3 9905 5115; fax: +61 3 9905 5499.
E-mail addresses: andreas.leibbrandt@monash.edu (A. Leibbrandt), raul.lopez@uam.es (R. López-Pérez).

[1] We say that a player C is a third party with respect to a player A if her material payoff does *not* depend on the decisions of A (note however that it is possible that the material payoff of A depends on the decisions of C; for instance, it could be the case that C sanctions A and thus reduces her material payoff). We also say that a player B is a second party with respect to A if her material payoff depends on A's decisions.

0167-2681/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jebo.2012.09.018

Despite their importance, there exist several uncertainties regarding how third parties sanction others. In particular, it is not clear whether third parties sanction in a different manner than second parties. On the one hand, third parties might sanction in a more impartial, "normative", and controlled manner, and less egocentrically (Fehr and Fischbacher, 2004). Adam Smith apparently had this idea in mind when he introduced the concept of the "impartial spectator" in his *Theory of Moral Sentiments*, a party who is not personally affected, making decisions from beyond the limitations of egocentric biases. In this line, the prevalence of institutions (like juries) that rely on third parties seems in accordance with the idea that third parties make more appropriate decisions. On the other hand, it also seems plausible that even third parties cannot completely eliminate egocentric biases (Ross et al., 1977; Babcock et al., 1995). In fact, the concerns about the selection of jury members in many law cases suggest that third parties can make very inappropriate decisions in the context of sanctioning (e.g. Kennedy, 1997).

In this paper we report the data from two experimental treatments, one on second party punishment, and the other on third party punishment. The two treatments include ten games which allow us to study and compare the motivations behind third and second party punishment. For this, we apply the classification method by El-Gamal and Grether (1995) and categorize subjects according to the theory that best explains their behavior in all ten games. By comparing the classification results for each treatment, we get some insights into the motivations of second and third parties. Our results also provide a picture of heterogeneity in agents' other-regarding preferences, another factor in which second and third parties might differ (e.g. the fraction of subjects who sanction in a normative, impartial manner could be different between third and second parties).

We consider many possible theories/motivations for third and/or second party punishment, including those advanced by recent theories of other-regarding preferences. Thus, theories of inequity-aversion like Fehr and Schmidt (1999) predict punishment of richer co-players if that reduces the payoff distance, while pure reciprocity theories (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Cox et al., 2007) are based on the idea that people harm those who harmed them. Further, Bolton and Ockenfels (2000) predict punishment of any co-player if that brings the aggressor's relative payoff closer to the average relative payoff, Levine (1998) posits the existence of spiteful types who punish indiscriminately and type-reciprocal agents who punish selfish or spiteful co-players, Falk and Fischbacher (2006) combine ideas from inequity-aversion and reciprocity, and López-Pérez (2008) predicts punishment of norm deviators.

Our study provides several insights about the differences and similarities in motivations behind third and second party punishment. To start, the classification analysis reveals two key non-selfish motivations behind the occurrence of both third and second party punishment. These motivations are two types of inequity-aversion for the third parties, and inequity-aversion and spitefulness for the second parties. If parsimony is the main goal, however, the Akaike Information Criterion suggests that a model including both inequity-averse and selfish agents (who never punish) is sufficient to explain *both* second and third party punishment. That is, in our games both types of punishment are predominantly targeted towards richer co-players, and as a result we do not find sharp differences in the motivations of both parties. We also observe that the strength of punishment depends heavily on the size of the payoff disadvantage for second parties and even more so for third parties.

Interestingly, our data shows that third parties often punish as intensely as second parties, which contrasts with previous findings (Fehr and Fischbacher, 2004), and that most third parties do not punish in a more normative and impartial manner than second parties. Therefore, our results suggest that second and third party punishment have more similarities than differences and show that models that include inequity-aversion as a motivation (specifically Fehr and Schmidt, 1999; Falk and Fischbacher, 2006) explain best the occurrence and strength of both types of punishment, whereas motivations like *pure* reciprocity or spite play a *relatively* minor role.[2]

The rest of the paper proceeds as follows. The next section compares our study with the related literature. Section 3 presents the experimental design and procedure. Section 4 reports the results from the classification analysis and analyzes which factors affect the occurrence and strength of second and third party punishment. The fifth section concludes.

2. Related literature

A large body of experimental research shows that subjects are often willing to spend money to reduce another player's payoff – i.e. to punish her – even if no future benefits can follow from this behavior. In the ultimatum game, responders frequently punish proposers for making unfair offers (Güth et al., 1982; Camerer and Thaler, 1995; Roth, 1995), while non-contributors are often punished in public goods games with a punishment stage (Ostrom et al., 1992; Fehr and Gächter, 2000). This literature has improved our understanding of the factors affecting punishment. For instance, the evidence from Anderson and Putterman (2006) and Carpenter (2007) indicates that punishment occurs less frequently as its price increases, and Nikiforakis (2008) shows that punishment in public goods games is frequently retaliated.

While the previous literature focuses on second party punishment, there is also a recent experimental literature on *third party* punishment (e.g. Zizzo, 2003; Fehr and Fischbacher, 2004; Carpenter and Matthews, 2009, 2012; Lieberman and Linke, 2007; Charness et al., 2008; Leibbrandt and López-Pérez, 2011). For instance, Lieberman and Linke (2007) show that the

[2] While *pure* reciprocity apparently plays a rather secondary role in our games, this is not to say that reciprocity is secondary in general. Indeed, the model by Falk and Fischbacher (2006) is in line with much of our evidence and it incorporates not only inequity-aversion, but also reciprocity.