Journal of Economic Psychology 67 (2018) 34–46

Contents lists available at ScienceDirect

Journal of Economic Psychology

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

Third-party punishment: Retribution or deterrence?

Fangfang Tan[a,*], Erte Xiao[b]

[a] Department of Public Economics, Max Planck Institute for Tax Law and Public Finance, Germany
[b] Department of Economics, Monash University, Australia

ARTICLE INFO

JEL classification:
C72
C92
D63
D70

Keywords:
Third-party punishment
Group decision-making
Retribution
Deterrence
Social dilemmas
Experiment

ABSTRACT

We conduct an experiment to examine the role of retribution and deterrence in motivating third-party punishment. Specifically, we examine how these two motives may play different roles depending on whether a third party is a group or an individual. In a one-shot prisoner's dilemma game with third-party punishment, we find that groups are more likely to punish when the penalty embeds deterrence rather than mere retribution. By contrast, when individual third parties make punishment decisions, they appear to give little weight to the deterrent effect of the punishment. In general, groups are less likely than individuals to impose punishment, even though the punishment is costless for third parties. However, decision-makers in the prisoner's dilemma game do not distinguish between an individual third party and a group third party.

1. Introduction

Punishment plays an important role in maintaining social order and economic relationships (for literature reviews see Chaudhuri, 2011; Gächter and Herrmann, 2009). Recent empirical findings on antisocial punishment and retaliatory punishment call into question the effectiveness of uncoordinated second-party punishment in sustaining social order. This research points instead to coordinated third-party punishment as an effective means of defending justice and avoiding escalation of violence that could lead to feuds.[1] In large societies, centralizing punishment in the hands of third parties is efficient. The reason is that the decisions of third parties are less selfishly motivated than second parties (Herrmann et al., 2008, Carpenter & Matthews, 2010; Jensen, 2010). While there is already a large literature devoted to understanding second-party punishment (e.g., Carpenter & Matthews, 2009; Casari & Luini, 2012; Xiao & Houser, 2005), more research is needed to improve our understanding of third-party punishment.[2]

In this paper, we examine what motivates third-party punishment. The literature on the philosophical, legal, and social reasons for

* Corresponding author.
E-mail addresses: fangfang.tan@tax.mpg.de (F. Tan), Erte.xiao@monash.edu (E. Xiao).

[1] Recent studies show that unrestricted second-party peer punishment can lead to antisocial punishments, feuds, or inequality (Cinyabuguma, Page, & Putterman, 2006; Denant-Boemont, Masclet, & Noussair, 2007; Herrmann, Thöni, & Gächter, 2008; Nikiforakis, 2008; Nikiforakis and Engelmann, 2011; Nikiforakis, Noussair, & Wilkening, 2012). These studies draw attention to the importance of designing restricted punishment mechanisms. In addition to the intervention of a third party, peer punishment can be regulated by predetermined rules (Andreoni & Gee, 2012; Xiao & Houser, 2011; Xiao & Kunreuther, 2015) or voting mechanisms (Ertan, Page, & Putterman, 2009; Kosfeld, Okada, & Riedl, 2009; Noussair & Tan, 2011; Tyran & Feld, 2006).

[2] Most of the third-party punishment literature focuses exclusively on individual third parties, see, e.g., Almenberg, Dreber, Apicella, & Rand, 2011; Carpenter & Matthews, 2010, 2012; Casari & Luini, 2012; Chavez & Bicchieri, 2013; Fehr & Fischbacher, 2004; Kurzban, DeScioli, & O'Brien, 2007; Marlowe et al., 2008; Ottone et al., 2008; Ouss and Peysakhovich, 2015; Walker & Halloran, 2004. We study both individual and group third parties. Some law literature (MacCoun & Kerr, 1988) studied mock jury decision-making and identified several biases in group judgment such as the defendant bias (being less willing to convict a defendant). Our paper focuses on uncovering the underlying motivation (regardless of whether certain behavior deserves punishment or not), rather than assessing the likelihood that a defendant committed a crime.

https://doi.org/10.1016/j.joep.2018.03.003
Received 9 January 2018; Received in revised form 17 March 2018; Accepted 17 March 2018
Available online 20 March 2018

Table 1
Payoff table of the prisoner's dilemma game.

		Person B	
		Option I(Cooperate)	Option II (Defect)
Person A	Option I (Cooperate)	30 / 30	40 / 15
	Option II (Defect)	40 / 15	20 / 20

Note: The exchange rate is 5 tokens to $1.

imposing punishment point out two motivations economists finds relevant: retribution and deterrence (Carlsmith, Darley, & Robinson, 2002; Durkheim, 1973; Kurzban & DeScioli, 2013; Woods, 2006).[3] Retribution is non-utilitarian, in that it is triggered by the desire to give wrongdoers what they deserve for the harm they inflict on others.[4] By contrast, deterrence is utilitarian, i.e., punishment is a tool used to prevent future norm violations. Differentiating between these two motives helps explain and predict the conditions under which people punish. Given that deterrence-driven punishment is correlated with the probability of future violation, it should not occur when future violation is no longer possible. By contrast, retributive punishment is generally correlated in some fashion with the harm the violator inflicted on the victims, and can occur even when the violator is no longer able to harm others. Understanding the role of these two motivations behind punishment decisions can facilitate designing regulations and rules. The reason is that people are more willing to comply with regulations and rules that are in line with their own judgment regarding what should be done (Tyler, 2006).

In naturally-occurring environments, third-party punishment is not limited to individual decisions. Rather, it is often decided by a group (e.g., a committee or a jury). Indeed, some have argued that coordinated group decision-making is one feature of punishment mechanisms in practice, due to the fact that the group input legitimizes the sanction (Boyd, Gintis, & Bowles, 2010; Ertan, et al., 2009; Guala, 2012; Ostrom, 2000). We thus investigate how each of the two motives comes into play when the third-party punishers are individuals and unitary groups respectively (i.e., a group that makes a joint decision). In summary, this paper attempts to answer two questions: (1) To what extent do third parties use punishment for retribution or to deter defectors?; and (2) How does the answer to the first question differ according to whether the third party is a group or an individual?

We study these questions using controlled laboratory experiments based on a framework that is of special interest to economists: a one-shot prisoner's dilemma (PD) game (see Table 1). The design of the punishment institution follows Tan and Xiao (2012). Players see their partner's decision before deciding whether to propose to punish their partner. The actual implementation of the punishment proposal depends on a third party's decision. If the third party approves the punishment, the punishment is implemented and the punishment proposer, rather than the third party, pays the cost of the punishment. If the third party rejects the punishment, the punishment is not implemented and no one's payoffs are changed.

As we elaborate later in detail, one advantage of this design is that it provides clean evidence for understanding the motivation behind punishment decisions. It also captures two features of coordinated punishment institutions in civilized societies. First, punishment often will not occur if the victims fail to make an accusation. Second, to ensure the justice and legitimization of punishment, the decisions of third parties (e.g., committee members, jury, or judges) often do not have any significant direct impact on their payoffs (Babcock & Loewenstein, 1997; Xiao, 2013; Guala, 2012). One example is the U.S. criminal system, in which a jury composed of people randomly selected from the eligible population has the opportunity to ascertain the guilt or innocence of an accused defendant.

The experiment has a two-by-two design, resulting in four treatments (see Table 2). A third party can either be an individual or a group (Individual vs. Group). The decisions of a third party can be announced either before or after the two players have decided whether or not to cooperate in a PD game and whether to propose punishment (Ex-ante vs. Ex-post). The treatment differences between Ex-ante and Ex-post measure the extent to which third-party punishment is motivated by deterrence.

Our main finding is that groups are more likely to punish instrumentally than individuals: Groups are more likely to approve punishment toward defectors when the punishment has a deterrence effect (in the Ex-ante treatment) than when the punishment can only be retributive (in the Ex-post treatment). By contrast, individual third parties punish in similar ways in both the Ex-ante and Ex-post treatments. Furthermore, consistent with previous studies like Tan and Xiao (2012), a significant amount of proposed punishment is not acted on by the third party. However, a new result of this study is that when the third party is a group, the punishment disapproval rate is even higher, despite the fact that punishing is a costless decision.

Our study contributes to the understanding of both third-party punishment behavior and group decision-making. First, our findings contribute to the recent new wave of theoretical and empirical work on coordinated third-party punishment mechanisms (see Guala, 2012 and discussion papers in that issue). Previous studies show that while third parties are willing to incur costs to

[3] There are many ways to classify the motives underlying third-party punishment behavior. For example, Kriss et al. (2016) examined the motives of punishment based on whether people intrinsically enjoy punishment or feel obligated to punish due to image concerns. In this paper, we classify the motives according to the goal of the punishment outcome: Is it to deter future violations or restore justice?

[4] Although retributive punishment is inconsistent with the homo economicus assumption, it could be driven by negative emotion towards norm violators (Casari & Luini, 2012; Hopfensitz & Reuben, 2009; Xiao & Houser, 2005), moral concerns (Cubitt, Drouvelis, Gächter, & Kabalin, 2011) or the relative earnings comparison (Bosch-Rosa, 2012; Dawes, Fowler, Johnson, McElreath, & Smirnov, 2007; Houser & Xiao, 2010). Numerous experimental studies show it occurs in social and economic relationships: Victims or even third parties often punish the wrongdoers in a one-shot scenario with no future interaction (Baldassarri & Grossman, 2011; Fehr & Fischbacher, 2004; Fehr & Gächter, 2000, 2002; Güth, 1995; Henrich et al., 2006, Kurzban et al., 2007).