Journal of Economic Psychology 31 (2010) 358–373

Contents lists available at ScienceDirect

Journal of Economic Psychology

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

Blind justice: An experimental analysis of random punishment in team production

Enrique Fatas [a,b,*], Antonio J. Morales [a,c], Paloma Ubeda [a]

[a] LINEEX, University of Valencia, Campus Tarongers, 46022 Valencia, Spain
[b] CBEES, University of Texas, Dallas, 800 West Campbell Road, Richardson, TX 75080, USA
[c] University of Malaga, Plaza El Ejido s/n, 29013 Málaga, Spain

ARTICLE INFO

Article history:
Received 15 February 2008
Received in revised form 23 October 2009
Accepted 18 January 2010
Available online 16 February 2010

JEL classification:
C92
H41

PsycINFO classification:
3020
2200

Keywords:
Public goods
Punishment
Experiments

ABSTRACT

We study the effect of blind punishment in a team production experiment, in which subjects choose non-observable effort levels. In this setting, a random exclusion mechanism is introduced, linked to the normalized group performance (R, from 0 to 1). Every round, each subject is non-excluded from the collective profit with probability R (and with probability $1 - R$ gets no benefit from the group account). Punishment does not depend on the individual behavior, but the probability of being punished reflects collective performance. As the exclusion probability is computed at the group level, no individual information is needed to implement exclusion. However, the probabilistic punishment risks to be perceived by subjects as procedurally unfair, as all subjects are treated in an identical, non-equitable manner (justice is blind). Our results suggest that random exclusion promotes a significant increase in cooperation. The effect seems to be associated with *hot* behavioral responses to punishment. However, convergence to full contribution is not observed.

1. Introduction

No group exists without norms, and norms are typically enforced by sanctions (see Posner & Rasmusen, 1999). This paper contributes to the experimental literature on social sanctions in social dilemmas, like public good games. This literature analyzes punishment in groups or teams, in which sanctions reduce both recipient and sender's earnings, and tend to be considered as altruistic or social as it is a second-order public good. Punishment was first analyzed experimentally by Yamagishi (1988) using a centralized mechanism, vertically enforced. Subjects contributed to a punishment account that was used to punish free riders. Ostrom, Walker, and Gardner (1992) first introduced non-centralized punishment in a common-pool resource setting and Fehr and Gächter (2000) in a public good experiment.[1] In these two papers punishment was carried out horizontally by individuals, not by a central authority. In all cases subjects were able to identify the full distribution of contributions in their group and punishment generated huge contribution gains.

* Corresponding author. Address: LINEEX, University of Valencia, Campus Tarongers, 46022 Valencia, Spain. Tel.: +34 961 625 161; fax: +34 961 625 415.
E-mail addresses: fatas@uv.es (E. Fatas), amorales@uma.es (A.J. Morales), paloma.ubeda@uv.es (P. Ubeda).

[1] The experimental literature has exponentially grown in the last years. Some papers are discussed below. Nikiforakis and Normann (2008) provide a brief survey of recent developments.

0167-4870/$ - see front matter
doi:10.1016/j.joep.2010.01.005

In this study, we specifically want to contribute to the understanding of enforcement mechanisms in teams when full information is not available.[2] In many real life interactions, individual information is hardly accessible, or simply too costly. However, punishment still takes place (and makes sense as it enhances cooperation). Managers sanction some members of defective teams without knowing whom to blame for the poor performance. Residents in one area may ostracize certain defecting neighbors, without having full information about who is individually responsible for the lack of cooperativeness. Teachers sanction some students in a rebellious class even when they cannot find someone accountable for the revolt. This mechanism is as old as the Roman army, in which generals (tribunes) punished one tenth of the soldiers in a legion to encourage discipline and fight cowardice.[3]

Note that in the examples above, managers, neighbors, teachers or tribunes choose in a random way, whom to punish. Justice is in this sense blind, as long as it cannot identify individual defectors. Some individuals are sanctioned, depending on facts of random nature (who is coming first to the managers' office, who is visiting the neighborhood, who is talking in the first row in class when the teacher turns around). All these examples have three additional interesting features. First, punishment is still social, in the sense that it pursues a collective goal. Second, not all team members are punished, because this is probably not necessary, maybe impossible or simply too costly. Third, even when sanctions are not linked to individual behavior, punishment still depends on the collective performance. Managers or teachers know the collective outcome and punish accordingly. So, the probability of being punished typically depends on the team's performance. Good teams (brave legions) are never punished.[4]

It is interesting to notice that randomness has already been considered from a theoretical point of view. Rasmusen (1987) proposes two contracts to achieve efficiency in teams: "massacre" and "scapegoat". In the former, all but one member (randomly chosen) of the team are penalized to pay a positive amount of money to the survivor, which in addition collects the joint product. In the latter, all but one (randomly chosen) shares the joint product and the penalty paid by the Guinean pig. Both mechanisms take advantage of risk aversion by fitting in random payoffs; i.e. players face a lottery in which with some positive probability they get negative payoffs.[5] A growing related literature has applied this idea to environmental issues.[6]

In this paper we study a punishment mechanism based on random exclusions. Exclusion is a rather common disciplinary measure in many real life situations. Shirking workers are fired (Shapiro & Stiglitz, 1984); uncooperative neighbors are not invited to social events; societal defectors are incarcerated or expelled (Hirshleifer & Rasmusen, 1989); and countries that violate international conventions are boycotted. In our design, every team member faces a common probability of being excluded from the team benefit, as high as one (as low as zero), when collective contribution to the joint outcome is 0% (100%) of their total endowment. Punishment realizations are i.i.d. across team members.

To test for the effectiveness of this random punishment mechanism, we run three different games in two alternative ways. A standard public good game based on the voluntary contribution mechanism (VCM) serves as a natural baseline. Relative to our benchmark treatment we test a random exclusion mechanism, with and without redistribution of the excluded share (as in Croson, Fatas, and Neugebauer (2006)). This design allows for a between subjects analysis across all three games. In addition, and to test for the role of random exclusion in overcoming cooperation failure,[7] the random punishment mechanism is introduced in some sessions after a common history of cooperation collapse (the VCM game). This allows for an additional within subjects analysis across games within the same sessions.

Croson et al. (2006) analyze a similar exclusion mechanism in different team production games.[8] In all games, punishment is deterministic, and based on competitive exclusion. The worst performer is excluded from the benefits of team production, so a competition between group members determines who is (not) going to be punished. Their experimental results show that excludability produces large increases in contribution. Note that even when full information about individual contributions is not needed to implement exclusion in this setting, an ordinal ranking of individual contributions is still necessary.

In our design, and contrary to Croson et al. (2006), exclusion is not based on competition. Some subjects are excluded from the collective benefit, but exclusion is not based on the relative performance of team members inside the team. Moreover, random punishment does not depend on the willingness to pay for punishing. This makes the success (or failure) of the mechanism independent of the existence of strong punishers. As it has been explained before, the *individual* information

---

[2] Fatas, Melendez-Jimenez, and Solaz (2009) analyze the role of incomplete information in public goods games with or without punishment in a variety of network structures. Their results suggest that the positive effect of punishment critically depends on the network completeness (that is, the existence of complete information).

[3] The Greek historian Polybius of Megalopolis described this *decimation* procedure in very interesting terms: 'The tribune assembles the legion, and (...) chooses by lots sometimes five, sometimes eight, sometimes twenty of the offenders, so adjusting the number thus chosen that they form as near as possible the tenth part of those guilty of cowardice. Those on whom the lot falls are bastinadoed mercilessly [...]. As therefore the danger and dread of drawing the fatal lot affects all equally, as it is uncertain on whom it will fall [...] this practice is best calculated (...) [to] inspire fear and correct the mischief' (cited in Hultsch, 1889).

[4] There is a large literature on the tension between identifying and punishing the offender in the economics theory of law enforcement, starting with Becker (1968), and the difficulties in identifying offenders and therefore letting the offend go unpunished. See Polinsky and Shavel (2000) for a survey and Miceli and Segerson (2007) for a discussion of several examples as well as a comparison between individual random punishment and group punishment.

[5] Theoretical research on efficiency properties of different mechanism in teams continued, as in McAfee and McMillan (1991) or more recently, Lazear (1998). Che and Yoo (2001) undertake the repeated version of the moral hazard problem in teams. Also, empirical papers appeared, i.e. Hamilton, Nickerson, and Owan (2003).

[6] Shortle and Horan (2001) review the literature on non-point source pollution. Segerson (1988) and Meran and Schwalbe (1987) independently apply Holmstrom's forcing contracts to non-point pollution settings. Xepapadeas (1991) proposes an alternative mechanism inspired on the asymmetric approach of Rasmusen's scapegoat. In this literature the mechanism is named random fine. He also proposes an alternative mechanism (group fine) based on subsidies and collective penalties, similar to the ones suggested by Meran and Schwalbe (1987) and Segerson (1988).

[7] Brandts, Cooper, and Fatas (2007) and Weber (2006) suggest that past experience critically determines current behaviour. Overcoming a cooperation failure becomes much harder; a strong test for mechanisms looking for efficiency.

[8] Margolis (2007) contains a lengthy discussion of the experimental data from Croson et al. (2006) and a theoretical model rationalizing them.