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.

Table 1
Treatments.

Treatment	Subjects	Groups	Game	
			Block 1	Block 2
I	28	7	VCM	VCM
II	48	12	VCM	RP
III	48	12	RP	–
IV	48	12	VCM	RPR
V	48	12	RPR	–
5	220	55		

requirements are kept at a minimum. Within each group, all members share the same probability of being punished, so strictly speaking, no individual information is needed for the mechanism to be implemented.

As far as we know, no experimental analysis of random punishment in teams has ever been done. An alternative random mechanism experimentally studied is random monitoring. Among many others, Nalbantian and Schotter (1997) experimentally tested it, reproducing the forcing contract proposed by Holmstrom (1982). The main differences between our design and random monitoring is that while individual information is obtained with random inspection, individual decisions are never used in our experiment. Moreover, sanctions coming from random monitoring are never random (cooperators are never punished). In addition, as Nalbantian and Schotter (1997) suggest, 'unless the probability of detection is great (and, therefore, costly to maintain), such monitoring schemes are likely to fail' (p. 316). In this sense, our random punishment is less demanding than random monitoring, from an informational point of view. We want to test whether it is more efficient.

We are aware of the practical limitations of this mechanism, mainly coming from its unfair nature. Regardless of their individual behavior, all individuals face the same probability of being sanctioned. As in any other public good game, free riders still get an equal share of the collective benefits (if they are not punished). In our experiment, defectors also generate a public bad: their low contributions produce a negative externality (random punishment). This unfair mechanism may deteriorate the 'motivational capital' of a firm, using the term of Akerlof and Kranton (2005). As Prendergast (1999) puts it, the success of organizations depends on members' willingness to take unselfish, efficiency enhancing actions. In this sense, blind punishment could easily harm the basis of public good provision in games: conditional cooperation.[9] Random exclusions may lower contributions to the public good by cooperative subjects, as they can now be excluded from the public good benefits.[10] So, we explicitly want to test whether the potential benefits of exclusions on contributions may disappear through the back door of a negative behavioral reaction to its blind, unfair nature.

Our results suggest that random punishment promotes efficiency in a significant way. Between subjects, random exclusion generates more public good provision (with and without redistribution). Within subjects, random exclusion survives a past experience of contribution failure, as a significant increase is always observed. These results cannot be explained by risk attitudes. A deeper behavioral analysis suggests that cooperative subjects negatively react to sanctions, when punished. However, conditional cooperation survives to the use of blind sanctions, generating a net positive effect.

The rest of the paper is as follows. Section 2 describes our experimental game, and provides a theoretical background. Section 3 analyzes decisions, earnings and behavioral patterns and Section 4 concludes.

2. Experimental design

Our experiment consists of three games and five treatments. Table 1 below contains a summary of all of them.

2.1. The games

Our first game is a standard public good game based on the voluntary contribution mechanism (VCM).[11] In the VCM, following Croson (2000) and Croson, Fatas, and Neugebauer (2005), in each period t, ($t = 1, 2, \ldots, T$), every individual i, $i = \{1, \ldots, n\}$, is endowed with an endowment in ECUs (experimental currency units), which can be either privately consumed (individual project) or invested in the public account (team project). The individual payoff to player i is

$$\pi_i^{\mathrm{VCM}}(c_1, \ldots, c_n) = e - c_i + \frac{b}{n}\sum_{j=1}^{n} c_j \tag{1}$$

[9] The existence of conditional cooperators is a well supported regularity. Subjects' contributions follow the contribution of the other subjects in the group (see, Fischbacher, Gächter, & Fehr, 2001). Some recent evidence suggests that punishment is a rich behavioral phenomenon, very sensitive to certain features of the game and the environment. Hermman, Gächter, and Thöni (2008) specifically document the existence of antisocial behavior triggered by punishment in different cultures; Nikiforakis (2008) shows that punishment is sensitive to retaliation; Houser, Xiao, McCabe, and Smith (2008) support the idea that the mere existence of a credible sanction crowds out norm-based social behavior and increase the likelihood of income maximizing behavior.

[10] An opposite effect is potentially possible: free riders may increase their contribution if they are reluctant to generate a public bad problem. Fatas and Godoy (2009) results suggest that this is highly unlikely.

[11] See Ledyard (1995).

where e denotes the endowment, c_i is player i's contribution to the group output and b/n is the marginal per capita return from the group contribution.

In the other two games, the VCM is complemented with a random punishment mechanism enforced at the individual level, i.e. with some probability, player i is excluded from the collective benefit, where probability realizations are i.i.d. across individuals. We assume that the non-exclusion probability $R(\cdot)$ is endogenous, as it is defined by the relative group contribution to the public good.[12]

$$R(c_1,\ldots,c_n)=\frac{\sum_{j=1}^{n}c_j}{ne} \tag{2}$$

A natural question arises as to the use of the collective benefit of excluded players. Following Croson et al. (2006), we consider two possibilities: in the random punishment with redistribution game (RPR) this benefit is redistributed among the non-excluded players, whereas in the random punishment game (RP), redistribution does not take place and the benefit is allocated outside the group (or kept by the firm).

2.2. Theoretical predictions

Under the standard assumption $0 < b/n < 1 < b$, each subject has a dominant strategy to free-ride on the contributions of others in the VCM game and that the unique SPE is socially inefficient. In the following we analyze the theoretical properties of the other two games under the assumption of risk neutrality.

In the random punishment game, player i's payoff is given by the following lottery

$$\pi_i^{\mathrm{RP}}(c_1,\ldots,c_n)=\begin{cases} e-c_i+\frac{b}{n}\sum\limits_{j=1}^{n}c_j & \text{with probability } R(c_1,\ldots,c_n)\\ e-c_i & \text{with probability } 1-R(c_1,\ldots,c_n)\end{cases} \tag{3}$$

Proposition 1 shows that this game shares with the VCM the fact that contributing nothing is Nash equilibrium, although a second Nash equilibrium emerges if the marginal per capita return of the public account is large enough. This new equilibrium entails full contribution by part of the players.

Proposition 1. In the random punishment game,

(i) $(0,\ldots,0)$ is Nash equilibrium.
(ii) $(e,\ldots,e)$ is Nash equilibrium if and only if $\frac{b}{n}\geqslant\frac{n}{2n-1}$.

Proof. See Appendix.

Note that the cut-off value for the marginal per capita return from the group contribution is decreasing in the group size n and that it converges to ½ as n goes to infinity.

For the random punishment game with redistribution, the lottery describing player i's payoff is more involved as the received payoff depends on the number of excluded players. Its definition makes heavy use of the binomial distribution.

$$\pi_i^{\mathrm{RPR}}(c_1,\ldots,c_n)=\begin{cases} e-c_i & \text{with probability } 1-R\\ e-c_i+b\sum\limits_{j=1}^{n}c_j & \text{with probability } R\binom{n-1}{0}R^0(1-R)^{n-1}\\ e-c_i+\frac{b}{2}\sum\limits_{j=1}^{n}c_j & \text{with probability } R\binom{n-1}{1}R^1(1-R)^{n-2}cdots\cdots\\ e-c_i+\frac{b}{k+1}\sum\limits_{j=1}^{n}c_j & \text{with probability } R\binom{n-1}{k}R^k(1-R)^{n-k-1}cdots\cdots\\ e-c_i+\frac{b}{n}\sum\limits_{j=1}^{n}c_j & \text{with probability } R\binom{n-1}{n-1}R^{n-1}0\end{cases} \tag{4}$$

In order to carry out our analysis, we restrict the analysis to a group size of four players which is actually the group size we use in the experiment.

Proposition 2. In the random punishment game with redistribution and for a group size of four,

(i) $(0,\ldots,0)$ is Nash equilibrium.
(ii) For values of b/n larger than 0.82, there exists a symmetric interior Nash equilibrium which converges to full contribution as b/n goes to 1.

[12] Different definitions of the probability of being punished give rise to different theoretical predictions and behavioural responses. In particular, an increase in the punishment probability should increase the willingness to contribute although it would make the mechanism more unfair for contributors. The investigation of alternative punishing probabilities is left for further research.

Proof. See Appendix.

The intuition behind Propositions 1 and 2 is that from the point of view of a player, and for given contributions of the remaining players, the corresponding uncertain prospect in the RPR game shares the same exclusion probabilities as in the RP game but incorporates larger payoffs because of the redistribution of the collective benefit of excluded players.

Hence, the incentives to shirk are larger when redistribution takes place explaining why a larger cut-off value for the marginal per capita return from the group contribution is required in the RPR game, 0.82, – in comparison to the RP game, 0.57 – for the existence of Nash equilibria with positive contributions to the public good.

2.3. Experimental design

In our experiment we use the following values: $e = 100$, $b = 2$ and $n = 4$, which imply a marginal per capita return from the group contribution of ½. Under these values, widely used in the literature (see for example Croson, 2000 and Croson et al., 2005), zero contribution is the unique Nash equilibrium of the three games.[13]

It can be shown that this theoretical result is robust to the presence of risk averse players. In both games, the cut-off value for the existence of other Nash equilibria is decreasing in the risk aversion parameter. The intuition is that the larger the b/n, the more attractive the public account. Hence, when b/n is large enough, full contribution becomes Nash equilibrium. However, shirking if the others are fully contributing is less attractive in the presence of risk aversion. Hence, the larger the risk aversion, the smaller the cut-off value of b/n needed.

Numerical computations using a constant relative risk aversion utility function for $b/n = ½$ show that full contribution is Nash equilibrium in the RP and RPR games for values of the risk aversion parameter larger than 0.93 and 1.85, respectively. According to Holt and Laury (2002), a small fraction of the population, around 13%, lies in the risk interval [0.68, 0.97] whereas a risk aversion parameter larger than 1.37 is labeled as "stay in bed".[14]

However, games RP and RPR might look quite different from a behavioral point of view. Note that although both procedures are unfair as exclusions from the public benefit are not tailored to fit individual misbehavior, further redistribution as carried out by the RPR procedure confers this procedure a more unfair character. Additional differences make necessary a polished design. In contrast to the RP procedure, redistribution makes the RPR mechanism be budget balanced, a property which might be valued by some but not definitively by the owner of the firm. Note that the owner is supposed to strictly prefer the RP mechanism because in principle it is cheaper. In addition, the literature warns us about the presence of history dependent subjects. In our context, it is worthwhile to analyze whether experimenting an episode of cooperation failure makes a difference in both mechanisms.

To address the above questions, a set of five treatments is implemented. Treatment I is our baseline treatment, in which participants face the VCM game described above, in two blocks of 20 rounds each (throughout the paper we denote a play of 20 periods of the same game as a block). In Treatments II and IV subjects play a VCM block and afterwards another block with the RP or RPR game, respectively. In Treatments I, II and IV subjects do not know at the beginning of the experiment that they may participate in another experiment. So, all these three treatments include a surprise restart between blocks.[15] In Treatments III and V, subjects only participate in a block facing the RP or the RPR games. Hence, this experimental design allows for both a between subjects comparison (e.g. across treatments for every block of 20 rounds) and a within subjects analysis (e.g. across blocks within each treatment).

In all treatments, participants were allocated into groups of four via a partners' random matching[16] (groups were randomly determined in the first period and remained unchanged through the experiment). The information available for the subjects is the ranked vector of contributions, so they could not trace back individual contribution over time.[17] Subjects received information about past individual contributions and earnings. They also got information about their group performance, whether or not they had been punished, and the number of excluded subjects in the group.

All z-Tree computerized sessions (Fischbacher, 2007) were run at the Laboratory for Research in Experimental Economics (LINEEX), at the University of Valencia. The experiment involved a total of 220 Economic and Business undergraduates without prior experience in similar experiments. Every subject participated in only one treatment. Participants were randomly allocated to private cubicles and instructions were read aloud.[18] Subjects filled out a questionnaire to check that they all

[13] According to Propositions 1 and 2, these results apply to smaller values of b/n, as for example 0.4, which is used by Fehr and Gächter (2000) and others. This is one of the reasons why we chose these parameters.

[14] Even if we assume that the 13% lies in the subinterval [0.93, 0.97], the probability that, in a group of 4 players, all of them have a risk parameter larger than 0.93 is $(0.13)^4$. This means that only one out of 3501 groups will consider full contribution as Nash equilibrium.

[15] This technique allows controlling for learning effects across time (Andreoni, 1988; Croson, 1996; Croson et al., 2005). No deception was involved, as earnings were fully independent and subjects had the chance to abandon the laboratory. We are pretty confident that subjects didn't feel forced to stay, even when in this paper all of them did. This has not been the case in other experiments ran in the same lab using the same technique; see Brandts, Fatas, and Lagos (2009).

[16] Some papers on VCM with punishment assume a strangers matching condition, e.g. Fehr and Gächter (2000), to avoid reputation effects. However, as Fehr and Gächter (2000) points out, the effect of punishment goes in the same direction when subjects are playing in the stranger-treatment as when they are playing in the partner-treatment.

[17] Fatas et al. (2009) has shown that no differences are observed in VCM when subjects are only provided aggregated information on group contribution.

[18] Instructions and pre-experimental quizzes, translated from Spanish, are in the appendix.

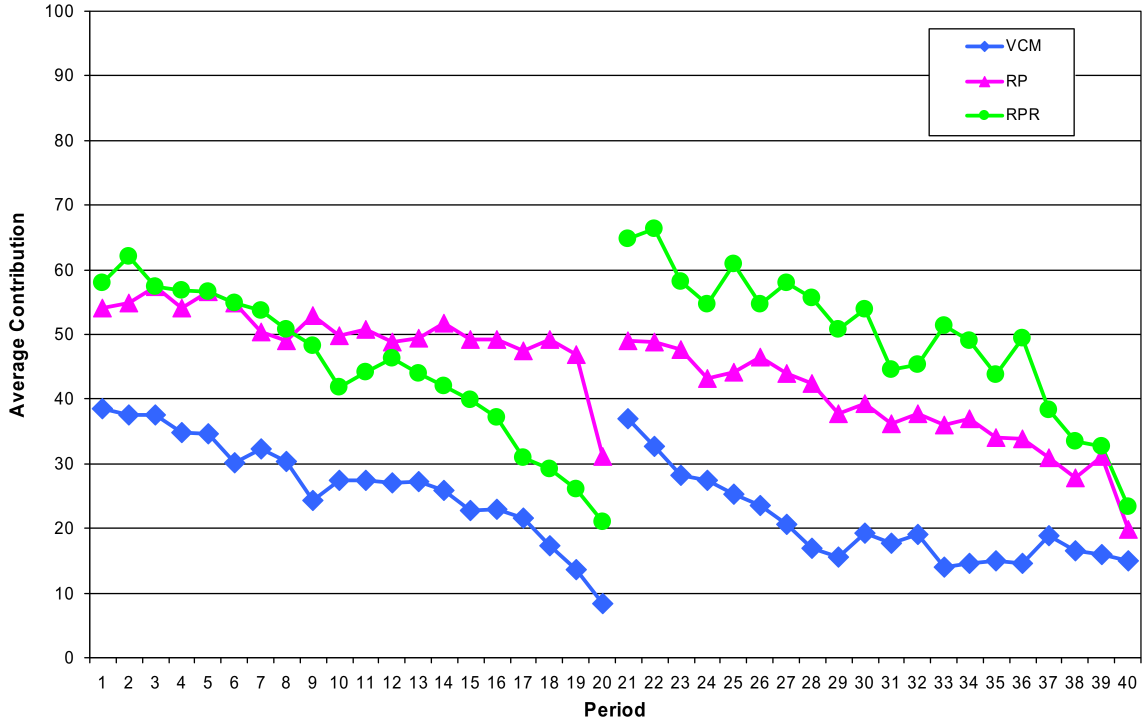

Fig. 1. Average contribution by game.

follow the basic logic of the game (that is, that they were able to compute payoffs). Average earnings were around 28€ and sessions lasted for 60–90 min, depending on the number of rounds.

3. Results

In this section we analyze contributions and earnings (as a proxy for efficiency), to conclude with a statistical analysis of the behavioral consequences of punishment at the individual level.

3.1. Contributions

Fig. 1 plots the average contribution to the public good (from 0 to 100, vertical axis), across games (VCM, RP or RPR) and blocks across rounds (from 1 to 40, horizontal axis). By simple inspection, contribution is higher in the two random punishment games (RP and RPR). Even more, for every round, game and block, average contribution is always smaller in the baseline game (VCM). However, contributions to the public good decline over time, in both blocks and games.[19] This decline is common in public good experiments based on a voluntary contribution mechanism (VCM) so our baseline game is very much in line with previous experimental studies.[20] However, this decline is usually not present in public goods with punishment between peers, as in Fehr and Gächter (2000). So, even when our random punishment mechanism promotes higher contribution to the public good, its effectiveness seems to be limited by its random implementation.

Using standard non-parametric analysis, we see that this negative trend is present in all blocks and games. Contributions are significantly lower in period 20 than in period 1,[21] and significantly lower in period 40 than in period 21.[22] A somehow different picture emerges when we consider the end-game effect and discard rounds 20 and 40. Contributions are still significantly lower in period 19 than in period 1 for VCM and RPR, and in period 39 than in period 21.[23] However, the decline is not significant in RP in the first block (p-value = .3078) and marginally significant in the second block (at the 10% level, p-value = .0545).[24]

[19] With the unique exception of the RP game in the first block. See Footnote 23.

[20] See the above mentioned survey by Ledyard (1995), or the more recent one by Andreoni and Vesterlund (2009).

[21] Wilcoxon sign-rank test at the group level yield the following p-values: .0000, .0280 and .0037 for the VCM, RP and RPR, respectively.

[22] The new p-values are .0280, .0053 and .0029 (again, for the VCM, RP and RPR games).

[23] Wilcoxon sign-rank test at the group level p-values are .0000 and .0042 in the first block and .0180 and .0150 in the second, for the VCM and RPR, respectively.

[24] We do not have a good explanation for this difference. Later in this section we show that no systematic differences are obtained between RP and RPR games. In addition, we show that individual behavioral reactions to punishment are very similar in both games. We can only conjecture that this decline is smaller because subjects perceive the random punishment mechanism without redistribution as less unfair.

Table 2
Average group contribution.

Game	Statistic	Block 1		Block 2	
		All rounds	Round 1	All rounds	Round 1
VCM	Average	27.07[a,b]	38.62[c,d]	20.38[e]	37.00[f,g]
	Std. dev.	30.87	30.42	19.23	26.29
	N (groups)	31	31	7	7
RP	Average	50.40[a]	54.06[c]	38.35	48.98[f,h]
	Std. dev.	32.82	26.88	35.96	29.76
	N (groups)	12	12	12	12
RPR	Average	45.04[b]	57.92[d]	49.47[e]	64.81[g,h]
	Std. dev.	33.78	30.36	42.01	38.39
	N (groups)	12	12	12	12

Mann–Whitney rank sum test *p*-values: a(.0041), b(.0027), c(.0011), d(.0005), e(.0225), f(.0912), g(.0018), h(.0285).

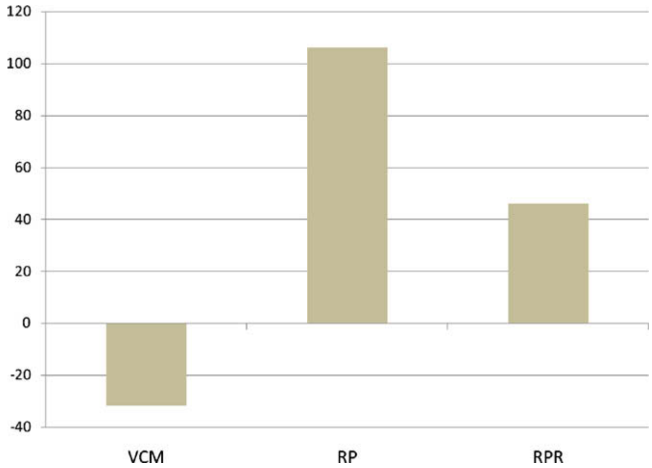

Fig. 2. Second block relative gains (within subjects' analysis).

Table 2 confirms the main findings of Fig. 1. Table 2 presents some descriptive statistics across treatments for both the first round, and the whole 20 rounds. We first carry out a between subjects analysis to establish some regularities about the existence of treatments effects. In period 1, contributions are significantly higher in RP and RPR than in VCM, in both blocks. While in our baseline game (VCM), subjects contribute 38.62% (37.00%) of their endowment in the first round of the first (second) block, contributions go up to 54.06% and 57.92% (48.98% and 64.81%) for the same rounds of RP and RPR games in the same block. So, the provision of the public good is roughly 50% larger whit random punishment. This increase is not only statistically significant, [25] but 'economically' relevant.

When we pay attention to the average contribution for all 20 rounds, the positive effect of punishment on contributions is even larger. Provision of the public good increased a 76% in the first 20 rounds, and a 115% in the second (averaging RP and RPR). This increase is significant in every comparison except in one: in the second block, for the comparison between RP and VCM.[26] This difference between games seems to suggest that subjects tend to contribute more in the RPR than in the RP, even when the decline is always larger. However, a comparison of contributions strongly suggests that differences between RP and RPR are never significant.[27] Moreover, our results suggest that the effect of our random punishment devices are similar, regardless of the block in which they are applied (contributions in the first block and the second block are not significantly different, within each game, RP or RPR).[28]

Treatments II and IV in our experimental design allow for an additional within subjects analysis. We can compare contributions to the public good of the same subjects with and without random punishment. In our view, it seems natural

[25] A Mann–Whitney rank sum test (at the group level) shows that in the first round of block 1, contributions are larger in RP and RPR than in VCM (*p*-values: .0011 and .0005). Same results in the first round of the second block (*p*-values: .0912 and .0018).

[26] A Mann–Whitney rank sum test (at the group level) shows that in block 1, contributions are larger in RP and RPR than in VCM (*p*-values: .0041 and .0027). In the second block, group contributions are significantly larger in RPR (*p*-value: .0225), but not in RP (*p*-value .2719).

[27] Mann–Whitney rank sum test (at the group level) shows no differences between RP and RPR in both the first and the second block (*p*-values = .4884 and .2727, respectively). If any, as the histograms in the appendix suggest, RP generates a strong polarization of provisions, making the provision of RP to be either very low or very high. This could also explain why the average contribution is sometimes flatter in the RP.

[28] A Mann–Whitney rank sum test (at the group level) shows no differences in the RP (*p*-value = .2727) and the RPR game (*p*-value = .5637), when comparing the average contribution in each block.

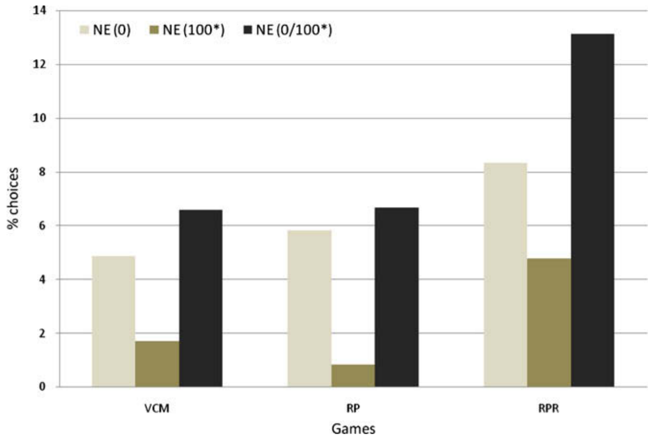

Fig. 3. Equilibrium play.

to check whether RP and RPR keep their positive effect after an experience of cooperation failure (playing a regular VCM in the first block). So, we compare contributions in the first block of Treatments II and IV (when all subjects play a VCM game) with the contributions of the same subjects in the second block (when they face a RP or a RPR game). Fig. 2 plots these changes across treatments presenting the difference between the second and the first average contribution, for the same subjects. Fig. 2 includes the same within subjects' analysis for the VCM (data coming exclusively from Treatment I).

The picture is clean. An expected decline occurs in the VCM, and contributions to the public good are lower in the second block by more than a 20%. However, the opposite result is observed in both RP and RPR. Contributions are more than doubled (an increase of more than a 100% is observed) in Treatment II, when RP is introduced, and the increase is above 40% in Treatment IV, when the random punishment mechanism includes redistribution (RPR). These differences are all statistically significant at the 5% level.[29]

Given the differences observed across treatments, an additional natural question arises: can the observed differences be explained by the existence of a new equilibrium? In Section 2 we showed that full contribution is an equilibrium strategy for extremely risk averse subjects (being matched with similar subjects). The data about the distribution of risk preferences in similar experiments suggested that the new equilibrium was statistically unlikely. But, as we did not elicit the risk preferences of participants, we cannot rule out this possibility and believe this question is worth analyzing.

A simple way to address this question is to look at the distribution of equilibrium choices to infer changes across games. Fig. 3 plots the proportion of subjects playing both the standard inefficient NE, in which nobody contributes to the public good (labeled 'NE(0)' in Fig. 3) and the efficient equilibrium in which everybody contributes, pooling across blocks ('NE(100)' in Fig. 3). Note that in both equilibria, no subject is ever actually punished. In the latter, because there is nothing to punish (there are no public good benefits); in the former, because the probability of being punished is zero (as $R = 1$). However, as we proved in Section 2, NE(100) is less attractive in RPR than in RP, because redistribution makes shirking more profitable (the collective profits of punished subjects is an additional payment).

Fig. 3 shows the percentage of equilibrium play (black bar) and how it is allocated between the efficient equilibrium (dark grey bar) and the inefficient equilibrium (light grey bar), across games. Even when the proportion of groups playing an equilibrium is relatively low in all treatments, the distribution of equilibrium play in RP looks almost identical to the distribution of equilibrium play in our baseline game (the VCM); we find no significant differences between both treatments for NE(0) and NE(100).[30] The same comparison generates a different result for RPR. Even when the new equilibrium is less attractive from a theoretical point of view, subjects play it more frequently and these differences are strongly significant.[31] We think that Fig. 3 convincingly supports the idea that positive results from random punishment do not derive from the existence of a new equilibrium based on risk aversion attitudes.

3.2. Earnings

The efficiency debate is always present in public goods experiments. Even when we do not present this mechanism as a realistic way to improve productivity in a real organization, the laboratory still allows for a clean analysis of this issue. In any

[29] Wilcoxon sign-rank test p-values at the group level are .0280, .0342 and .0281 for the VCM, RP and RPR, respectively. In all three cases, we match absolute group contribution levels in the public good for the same participants, in the first and second block. In the VCM, we need subjects to play the same game in both blocks, so only decisions coming from Treatment I are considered. From Fig. 2, we see that the difference is negative in Treatment I (when they play a VCM in the second block) and positive in Treatments II and IV.

[30] Chi-square test p-values are .904 (.137) when comparing NE-0 (NE-100) in RP relative to VCM.

[31] Chi-square test p-values are .000 (.000) when comparing NE-0 (NE-100) in RPR relative to VCM. These differences are mostly due to the presence of a group that fully contributes all the way long from the very beginning.

public good experiment, the provision of the public good is a sufficient measure of the efficiency achieved by subjects under a specific institution. When any costly sanctioning mechanism is considered, both the benefits of a larger contribution to the public good and the punishment costs must be evaluated to compute the net effect of punishment.

Fig. 1 and Table 2 made a clear point about the positive effect of punishment in contributions. Fig. 4 incorporates punishment costs (exclusion, in our case) to the picture and compares the net effect, presenting the total earnings of subjects participating in different games. Earnings are a commonly used methodology (beginning with Fehr & Gächter, 2000) to assess the net effect of punishment on efficiency. One of the critical assumptions in experimental economics is that subjects are intrinsically motivated by earnings in the laboratory.[32] Fig. 4 plots earnings using a 2000–4000 scale, because these two quantities correspond to the payoffs associated to zero and full contribution symmetric profiles. Fig. 4 suggests that relative to our baseline (VCM), RP reduces subjects' earnings in both blocks.[33] Contributions gains are not enough to compensate punishment costs. However, RPR makes subjects better off. This relative difference is not unexpected, because punishment in RPR becomes neutral for earnings as penalties are redistributed. Our data suggest that, both in the first and the second block, subjects make significantly more money in RPR than in VCM.[34]

Our results are very much in line with previous findings of the experimental literature on punishment (see Nikiforakis, 2008; Ostrom et al., 1992; Sefton, Shupp, & Walker, 2005). In most of the public goods games with punishment, earnings are not higher (or significantly lower) when punishment between peers is implemented. However, we believe that the interpretation of our results needs to be especially cautious. Relative to other studies, punishment is institutional in the sense that it is vertically implemented. In our experiments, participants cannot punish other subjects in their group. This makes punishment no costly at all in the presence of redistribution and transforms earnings into only a partial proxy for efficiency.

If we interpret these random punishment devices as a compensation scheme to increase effort levels in team production, earnings become salaries. They are still a very reasonable way to measure efficiency for employees (our experimental subjects), but not for the employer. In this view, redistributing is a costly option for firms given that they pay as a bonus what they have collected using the random punishment mechanism. In this sense, we can conjecture that random punishment with redistribution is not characterized as efficient by our data, as the same effort levels are obtained without redistribution. Given that there are no firms in our experimental setting, it is not feasible to assess efficiency for employers in a further, more specific way.

3.3. Behavioral determinants

We want to finish this section understanding what happened in the laboratory. Random punishment generates more contribution, not always more earnings, even when it cannot stop the usual decline in public good provision. Risk aversion does not appear as a consistent explanation for these results. Then, why did this happen? How did our participants react to a blind punishment? Given that our mechanism does not distinguish between those who contributed a lot and those who did contribute nothing to the public good, did blind punishment generate a different reaction in 'good' and 'bad' performers?

Fig. 5 tries to answer this general question presenting a dispersion figure of current and past contributions to the public good (vertical and horizontal axis, respectively). As our purpose is to identify the effect of random punishment on changes in contribution, Fig. 5a distinguishes between those punished in the previous round ($t-1$, blue solid diamonds) and those subjects who were not (red hollow circles). To make the analysis more salient, we also include a trend line (a linear estimation of the relationship) with the same colors (solid blue for punished subjects, dashed red for unpunished participants).

Fig. 5a pools data from all blocks and periods. It generates a very informative picture about the dynamic of contribution changes in the experiment. In other words, subjects' lagged contributions are critical to understand the role of random punishment. Unpunished subjects (red dots and red line) behave differently than punished subjects (blue dots and blue line). The difference is especially salient in the right side of Fig. 5a. While low contributors in the previous round do not react to punishment (relative to those unpunished, the red and the blue lines are really close to each other in the left side of the *X*-axis), high contributors react negatively when the random mechanism punished them (the blue line lies well below the red one in the right side of the *X*-axis). This difference is consistent with the idea that when subjects are contributing most of their endowment to the public good, they decrease their contribution when punished.

This behavioral reaction is not only consistent with the blind nature of our mechanism, but with the fact that random punishment has a positive net effect on contributions. In the games with random punishment (the RP or the RPR), some punished subjects adjust downwards their contribution relative to those participants who are not punished. But, random punishment makes all subjects to contribute more (overall, on average) than in our baseline game (the VCM).[35]

Recall that Fig. 5a pools data from RP and RPR (from all groups, rounds and blocks) and displays current and past contribution levels in absolute terms. Given that our experiment produces very different, sometimes polarized, group dynamics, a

[32] That is, the 'induced value theory', as called by Smith (1976). Please note that this does not preclude about their motivations. Subjects can be selfish in the sense that they are exclusively motivated by their own earnings or have any kind of social preferences.

[33] Mann–Whitney rank sum test p-values are .000 and .001, respectively, at the group level.

[34] Mann-Whitney rank sum test p-values are now .0304 and .0081.

[35] Fig. 5a (and 5b and 5c) also conveys a different message, not specific to our punishment mechanism but to any public good experiment: conditional cooperation. Large (small) contributions in a round are followed by smaller (larger) contributions in the next period. This conditional cooperation will be supported by additional analysis later in this section, and it is consistent with the fact that most dots are below the bisectrix in the right side of the figure (top contributors decrease their contributions), and most dots are above it in the left side (bottom contributors increase their contribution).

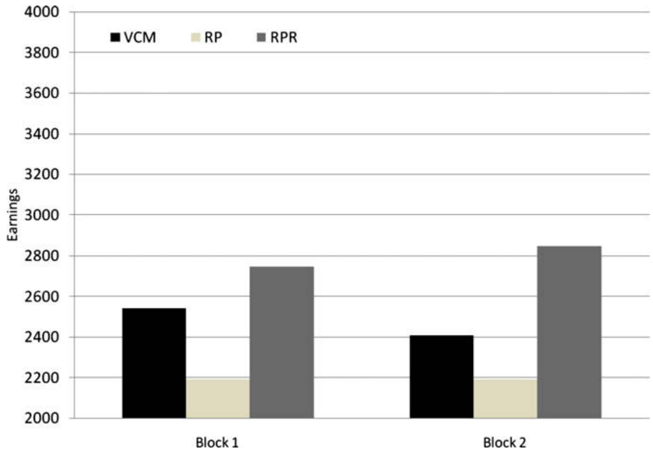

Fig. 4. Earnings.

sensible question is to investigate to what extent the different group trends blur the general result presented by Fig. 5a. Fig. 5b and c explicitly address this issue disentangling between the contribution adjustment of good and bad performers in every group.

Fig. 5b (5c) plots data of the best (worst) performers in each group: a subject is classified as good (bad) performed if his contribution in the previous round was above (below) her group median contribution. Fig. 5b and c supports the idea that it is subjects' relative contribution, rather than their absolute performance, the determinant of their behavioral reaction to punishment. While in Fig. 5b, regardless of the absolute contribution level, punishment does not make a difference (blue and red lines are very close to each other), this is not the message conveyed by Fig. 5c: good performers decrease their contribution when punished (blue dots and blue line, relative to unpunished participants), unless their absolute contribution level is very low.[36]

Fig. 5 hence gives us a descriptive explanation of the effect of punishment in contributions changes. But it tells us very little about the significance of this effect, or the convenience of pooling data across games and periods, as some of the results presented earlier in this section suggest that some differences may exist between RP and RPR.[37] To accomplish this goal, we need to descend from the robust non-parametric analysis to the more informative arena of panel data estimations.[38] This methodology allows us to get a deeper insight into the individual determinants of behavior, and understand how different subjects react to punishment in different treatments.

Table 3 presents the estimations of three models, using a panel data technique (across subjects and rounds) with random effects at the individual level (to capture idiosyncratic behavior). Given that we want to understand the behavioral reaction to punishment, we focus on games in which random punishment is present and discard observations in the baseline game. We also cluster standard errors at the group level to control for the fact that our subjects are repeatedly interacting in groups, generating statistically non-independent individual observations.

The dependent variable in all these models is the contribution in period t (c_{it}). The list of explanatory variables includes an intercept (*Constant*), the round (*Period*), a dummy variable that takes value 1 only when the subject was punished in the previous round (*LagPUN*), or two periods ago (*Lag2PUN*), and a dummy capturing the relative performance of subjects in their group in the previous round (relative to the median, *LagDRP*). Finally, *LagDRP* ∗ *PUN* is the interaction term between *LagDRP* and *LagPUN*.

To control for the different treatments and games, the dummy *Redistribution* (*Block*) takes value 1 in the RPR game (in the second block, 0 for RP). Given the strong impact of punishment in the first round of every block suggested by some previous results, we control the levels of contribution in the first period with *Contri_1* (the absolute individual contribution to the public good in round 1, in every block). This gives us a chance not only to control for the heterogeneity of subjects in period 1, but to differentiate between the anticipatory effect of random punishment in different games (RP vs. RPR) and its dynamic effect.

The last set of independent variables incorporates conditional cooperation patterns. The evidence about conditional cooperation in public good settings is overwhelming (see Croson et al., 2005; Fischbacher & Gächter, forthcoming; Fischbacher

[36] Note that the probability of being punished decreases with the group performance, so the distribution of red points (unpunished subjects) tends to be naturally biased to the right (higher absolute levels of contribution to the public good), relative to the distribution of punished subjects (biased to the left). This comes naturally from the fact that when you contribute less, you are more frequently punished.

[37] See the previous footnote for another source of spurious and blurring effect.

[38] Panel data estimations have been widely used in the experimental literature in general, and in the analysis of public goods experiments in particular, with and without punishment, see Brandts et al. (2007), Frechette (2010), Frechette, Kagel, and Morelli (2005), Burgess and Pande (2005) for some examples of this analysis. Woolridge (2003) discusses the limitations of this methodology.

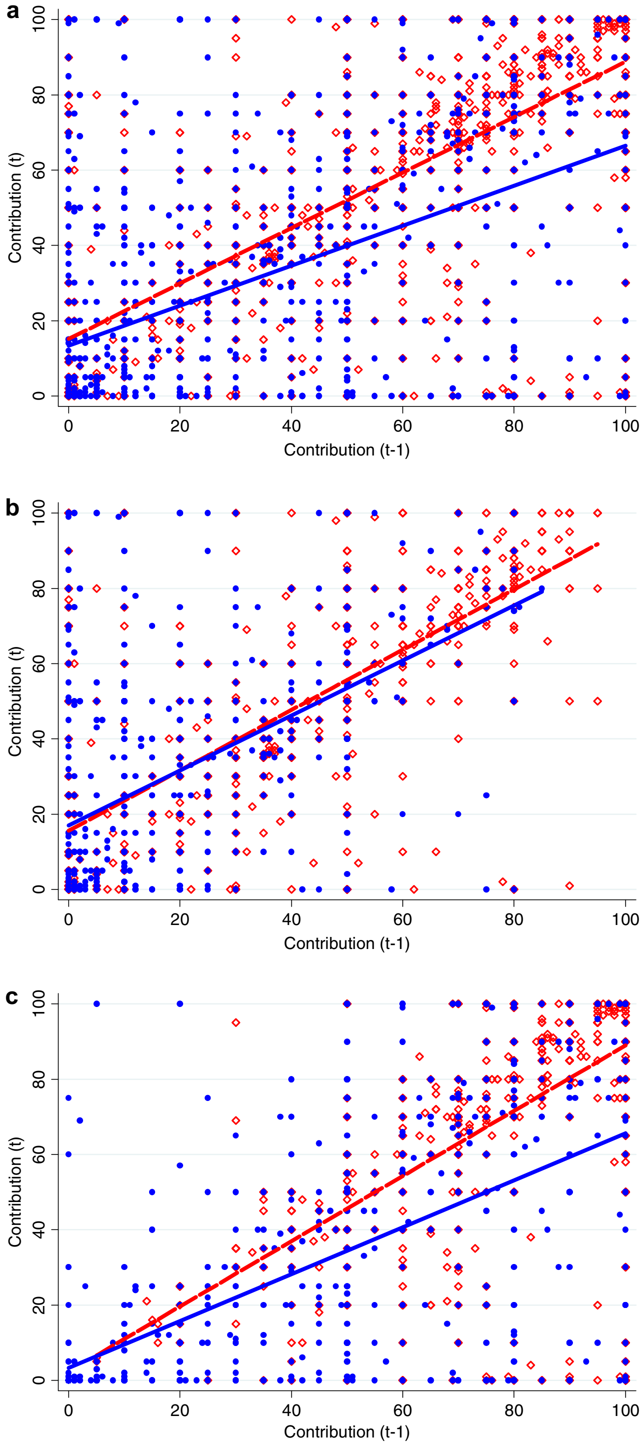

Fig. 5. Contribution adjustment in games RP + RPR. Dashed red: non-punished in t − 1. Solid blue: punished in t − 1. a) All subjects, b) Bottom performers (lagdrp = 0), c) Top performers (lagdrp = 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

et al., 2001). This evidence suggests that conditional cooperation is a solid behavioral pattern in public goods experiments. Given the group size in our experiments, we can easily look at the lagged maximum, *LagMax(−i)*, median, *LagMed(−i),* and minimum, *LagMin(−i)*, contribution of the other three members of the group. We are interested in knowing whether conditional contribution pattern is altered by the inclusion of a random punishment and, at the same time, in disentangling between relative performance and conditional cooperation effects.

Table 3
Individual behavior RP and RPR data.

	Dependent variable: contribution		
	(1)	(2)	(3)
Constant	26.997*** (5.858)	9.942** (3.888)	6.554* (3.928)
Period	−.655*** (.129)	−.545*** (.100)	−.541*** (.095)
LagPUN	−5.779*** (1.038)	−4.958*** (1.094)	2.494* (1.371)
Lag2PUN	−4.813*** (1.297)	–	–
LagDRP	–	14.580*** (1.177)	22.155*** (1.922)
LagDRP∗PUN	–	–	−13.910*** (1.992)
Redistribution	−2.150 (2.475)	−2.538 (1.720)	−2.211 (1.704)
Block	−1.830 (2.473)	−2.160 (1.679)	−2.314 (1.668)
Contri_1	.201*** (.034)	.150*** (.026)	.143*** (.025)
LagMax(−i)	.252*** (.032)	.349*** (.033)	.336*** (.031)
LagMed(−i)	.097** (.041)	.111*** (.041)	.107*** (.041)
LagMin(−i)	.163*** (.040)	.225*** (.038)	.234*** (.038)
R^2 within	.178	.203	.216
R^2 between	.724	.835	.840
R^2 overall	.462	.519	.528
Prob > chi^2	.000	.000	.000
N	3456	3648	3648
Clusters	48	48	48

* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table 3 above serves several purposes. First, it strengthens the results of the non-parametric analysis performed in previous sections. In this sense, in line with the decline suggested by Fig. 1, supported via a pair-wise comparison of average contributions in rounds 1 and 20 (or 19), the variable *Period* is always significant and negative. In line with our non-parametric analysis, *Redistribution* and *Block* are never significant. On one hand, this implies that no significant differences are observed between games RP and RPR, in line with most of the evidence analyzed in Table 2. On the other hand, all coefficients for *Block* are not significantly different from zero, supporting the idea that the effect of random punishment is independent of the block in which it is applied (after playing a VCM game for 20 rounds, or from the very first round of the game).

Second, Table 3 uncovers the motivation underlying individual decisions. *Contri_1* coefficient is always significant and positive, suggesting that subjects anticipate the punishment effect on behavior and choose higher contributions from the very beginning. Interestingly, the inclusion of this variable in the estimations allows controlling for different initial contribution levels. In addition, all lagged individual contributions play a positive role in the three models: subjects adjust their contributions to the observed contributions of others, for good or for bad. This result is not only in line with the above mentioned experimental analysis of similar games with conditional contributors. It additionally suggests that conditional cooperation is not destroyed by the randomness of the sanctioning mechanism. We conjecture that this supports the idea that the perverse effects of a blind mechanism are not strong enough to cancel conditional behavior among participants in our public goods games.

And third, Table 3 allows the analysis of punishment's effects. Individual contributions are better understood when we incorporate the individual effect of punishment. Its net effect is the one anticipated by Fig. 5. Model I shows that random punishment has a rather persistent negative effect in contributions. That is, the overall effect of random punishment in contributions is negative in the sense that, relative to those subjects in RP and RPR who were not punished in the previous round, punished subjects contribute less.

More interestingly, punishment is strongly mediated by the relative performance of subjects. Top performers (subjects with contribution levels above the group median in the previous round) keep contributing more than the others (as suggested by the positive, and huge, coefficient of *LagDRP* in model II). However, model III introduces a clarifying interaction term ($LagDRP * PUN$). Past top performers diminishes their contribution when punished (as suggested by the negative coefficient of the interaction term), while keep contributing more than the rest when not punished (as suggested by the positive coefficient of *LagDRP*). The net effect of punishment for the rest of the punished subjects (the bottom performers) is now positive (even when smaller and marginally significant).

Table 3 shows a clear individual contribution pattern. A heterogeneous population still governed by conditional cooperators (the proportion of conditional cooperators is high enough to make the three coefficients of the conditional cooperation variables positive and significant), reacts in a hot behavioral way to random punishment. Top performers clearly decrease their contribution when punished, while other subjects react marginally, but do not decrease their contribution.[39] All these results survive when we control for game, order and subjects types.[40]

[39] As a referee pointed, there are two reasons why top performers reacted differently: because they are punished per se or because the other players in their group did not contribute. We cannot discriminate between these two.
[40] Even to the specification of the model, see additional estimations in the appendix.

4. Conclusions

Groups sanction defectors. Sometimes they do it on an individual basis, making free riders to contribute, as Croson et al. (2006) shows. Sometimes individual information is not available but individuals still punish members depending on their collective performance. In this paper, we have analyzed the impact of random punishment in an experimental public good setting. Each round, a member of a group receives her share of the public good with a probability that matches the relative aggregate contribution to the public good.

The main advantage of random punishment relative to other punishment schemes considered in the literature is that the information requirements are low: no information at individual level is required at all. Hence, a mechanism of this sort seems especially applicable in situations where only information on the group level is available.

In our experiments, random sanctions do not eliminate the usual contribution decline, but contributions are significantly larger than in a baseline without any random sanction. Moreover, this result is independent of the way the excluded share is distributed (or not distributed at all) among group members. However, the minimal informational requirements impose a kind of a compromise between information (the input feeding random exclusion) and efficiency (the output) of the mechanism. In other words, we minimize the informational requirements at a cost. Free riders and co-operators within the same group will be treated equally, and punished with the same probability. It seems natural to consider that the blindness nature of the mechanism might jeopardize the basis of public good provision in games: conditional cooperation.

A conditional co-operator contributes to the public good as long as she observes others contributing. In a public good game with horizontal punishment, conditional co-operators may punish free riders to eliminate the advantages of defection. In a public good game with blind sanctions, punishment is not driven by individual behaviour but by collective performance. So, random exclusions might lower contributions to the public good by cooperative subjects if they anticipate that both co-operators and free riders will be excluded from the public goods benefits with the same common probability. Our experimental results support the positive rationale of using this unfair rule, in the sense that even when top contributors negatively react to being unfairly treated, punishment generates significant contributions gains.

We want to finish the paper with the following straightforward interpretation of our findings: blind justice is better than no justice at all. Conditional co-operators survive the perverse effects of random punishment and are able to generate large contribution gains, even when defectors do not take the hint when punished.

Acknowledgement

We thank the editor and two anonymous referees for valuable comments. Antonio Morales and Enrique Fatas acknowledges the financial support received from the Higher Council of Innovation, Science and Education of the Andalusian Regional Government through the Excellent Research Project number SEJ-1586 entitled "Economía Experimental". Enrique Fatas acknowledges the financial support received from the Spanish Ministry of Science and Innovation, Grants SEJ2007-66581 and ECO2008-04784, and the Valencia Regional Government, Prometeo Research Grant GRIECES.

Appendix A. Proofs

Proof of Proposition 1. We divide the proof in three steps.

Step 1. We first show that the RP Game has no interior Nash equilibrium. Note that it will suffice to show that the expected payoffs function is a convex function of the own contribution and therefore that any stationary interior profile of contributions would a minimum rather than a maximum of the function.

$$\frac{\partial^2 E(\pi_i^{\mathrm{RP}})}{\partial c_i^2} = 2be\left(\frac{\partial R(\cdot)}{\partial c_i}\right)^2 = \frac{2b}{en^2} > 0 \tag{5}$$

As we see, the second derivative is positive for any contribution profile.

Step 2. We next show that (0, …, 0) is Nash Equilibrium of the RP Game. We need to prove that contributing nothing is best response to a zero contribution by the rest of the group. To this end, note that in this case the non-exclusion probability $R(\cdot)$ cannot be greater than a half, as we easily show

$$R(0,\ldots,c_i,\ldots,0) = \frac{1}{en}\sum_{j=1}^{n} c_j = \frac{c_i}{e}\frac{1}{n} \leqslant \frac{1}{n} \tag{6}$$

The first derivative of expected payoff is

$$\frac{\partial^2 E(\pi_i^{\mathrm{RP}}(0,\ldots,c_i,\ldots,0))}{\partial c_i} = -1 + 2R(\cdot)\frac{b}{n} \leqslant -1 + \frac{b}{n} < 0 \tag{7}$$

As we see, the first derivative is strictly negative, meaning that it is in the interest of player *i* to lower his contribution till reaching zero contribution.

Step 3. We finally show that $(e, \ldots, e)$ is Nash equilibrium of the RP game if and only if $\frac{b}{n} \geqslant \frac{n}{2n-1}$. Because expected payoffs are a convex function of the own contribution, it is the case that the best response to a given contribution of the rest of the group is one of the two corners of the strategy space, either $c_i = 0$ or $c_i = e$. This implies that $(e, \ldots, e)$ is Nash equilibrium if and only if $E(\pi_i^{\mathrm{RP}}(e, \ldots, e)) \geqslant E(\pi_i^{RP}(e, \ldots, 0, \ldots, e))$. We now prove that this is the case:

$$E(\pi_i^{RP}(e, \ldots, e)) \geqslant E(\pi_i^{RP}(e, \ldots, 0, \ldots, e))$$
$$\Updownarrow$$
$$Be \geqslant \frac{1}{n}e + \frac{n-1}{n}\left(e + \frac{n-1}{n}be\right) \tag{8}$$
$$\Updownarrow$$
$$\frac{n}{2n-1} \leqslant \frac{b}{n}$$

This last step completes the proof of Proposition 1.

Proof of Proposition 2. We divide the proof in two steps.

Step 1. $(0, \ldots, 0)$ is Nash equilibrium of the RPR game. We prove that if the rest of the players are contributing zero, then it is in the interest of a single player to also contribute nothing. Recall that by contributing nothing, player *i*'s payoff is *e*. By contributing a positive amount $c_i > 0$ to the public good, player *i*'s expected payoffs can be written as follows $E(\pi_i^{\mathrm{RPR}}(0, \ldots, c_i, \ldots, 0)) = e - c_i((b/4)R^2(2 + 4(1-R) + R^2) + (1 - bR))$. Taking into account that $E(\pi_i^{\mathrm{RPR}}(0, \ldots, 0)) = e$, that $2 + 4(1-R) + R^2 > 0$ and that $bR \leqslant 1$ it follows that $E(\pi_i^{\mathrm{RPR}}(0, \ldots, c_i, \ldots, 0)) < E(\pi_i^{\mathrm{RPR}}(0, \ldots, 0))$ as we aimed at.

Step 2. The first derivative of the expected payoff with respect to own contribution is $\partial E(\pi_i)/\partial c_i = -1 + (b/4)R(8 - 18R + 16R^2 - 5R^3)$. Noting that $b/4 < 1$ and that $R(8 - 18R + 16R^2 - 5R^3)$ reaches a maximum value of 1.216 for $R = 2/5$, the derivative $\partial E(\pi_i)/\partial c_i$ is negative for values of *b/4* smaller than $1/1.216 \approx 0.82$. For values larger than 0.82, the first order condition has two stationary points on R: the first one is smaller than 2/5 while that other one is larger than 2/5. The second derivative $\partial^2 E(\pi_i)/\partial c_i^2 = b(2 - 5R)\,(1-R)^2$ confirms that the stationary point larger than 2/5 is a maximum. It is easy to see that this maximum converges to 1 as *b/4* goes to 1.

Appendix B. Instructions and test (RP treatment)

B.1. Instructions

The aim of the experiment is to study how individuals make decisions in some environments. Instructions are easy and you can make a non-negligible amount of money if you follow them carefully. Money will be privately paid at the end of the experiment. Should you have any questions please raise your hand before asking them. Any communication between you and other participants is strictly forbidden. If you do not follow this rule, you will be excluded from the experiment.

1. The experiment consists of 20 rounds. In each round you are member of a group of four participants. The composition of each group is randomly determined at the beginning of the experiment and does not change along the experiment. You will never know the identities of the other group members.
2. In each round every participant gets an initial endowment of 100 ECU (Experimental Currency Units). You must only decide how much of this amount you want to assign to a Group Project. The remainder will automatically be assigned to an Individual Project.
3. Your payoff from the Individual Project (IP) equals your assignment to the Individual Project and does not depend on the decisions of others.
4. The payoff from the Group Project (GP) depends on the total amount of ECU assigned to this Project. That is, the sum of your assignment to the Group Project and the corresponding assignments of the other three members of your group. This amount will be doubled and equally divided among the four group members.
5. The benefit arising from GP is determined as follows. The computer computes the ratio R:

$$\mathrm{R} = \frac{\text{Group Allocation to GP}}{\text{Maximum Allocation}} \times 100 = \frac{\text{Group Allocation to GP}}{400} \times 100$$

6. This ratio determines the probability that a group member receives the payoff from GP. That is, a subject gets the GP benefit with probability R, and gets nothing with probability $(1 - R)$. An example will help you understand the rule. Assume that R is 0.7; then each member of the group has a 70% chance of earning the GP benefit, and a 30% chance of getting nothing from GP.
7. In summary:

$$\text{Payoff} = \begin{cases} & \text{Prob} & \text{IP profit} & \text{GP profit} \\ \text{I am not excluded from GP :} & R & (100\text{ ECU} - \text{my IP allocation}) & + \frac{2\times\text{GP}}{4} \\ \text{I am excluded from GP :} & (1-R) & (100\text{ ECU} - \text{my IP allocation}) & - \end{cases}$$

   After each round you will get information about individual allocations to the GP in your group, ranked from top to bottom, so you will not be able to track individual decisions across periods. You will also get information about your payoffs (both GP and IP profits) and the R value. A table will record all available information about past rounds.
8. At the end of the experiment, the sum of your individual payoffs over the 20 rounds will be privately paid to you at the exchange rate of 200 ECU = 1€.

B.2. Test

Choose the correct answer and fill out the gaps. When you finish, raise your hand and an assistant will check the answers. The assistant will inform you whether your answers are correct or not. If you have made any mistake the assistant will give back the test and you will have to fill it again. You cannot ask any question to the assistant. Any communication between you and other participants is forbidden. If you do not follow this rule, you will be excluded from the experiment.

1. My benefit depends on other group members decisions.

   ☐ True ☐ False

2. In order to calculate my benefits in one period I don't need to care about any data from other periods and other group members.

   ☐ True ☐ False

3. The composition of each Group does not change along the experiment.

   ☐ True ☐ False

4. The following table shows some individual decisions in a group. Fill out the gaps.

Participant	Assign to Group Project (GP)	Assign to Individual Project (IP)	Group Allocation to GP	Benefit from IP	Benefit from GP	Total benefit
1	0					
2	25					
3	75					
4	100					

Participant	Assign to Group Project (GP)	Assign to Individual Project (IP)	Group Allocation to GP	Benefit from IP	Benefit from GP	Total benefit
1	0					
2	0					
3	100					
4	100					

5. The highest individual payoff in one period is 200 ECU

   ☐ True ☐ False

6. The lowest individual payoff in one period is 50 ECU

   ☐ True ☐ False

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.joep.2010.01.005.

References

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the economics of organizations. *The Journal of Economic Perspectives, 19*(1), 9–32.

Andreoni, J. (1988). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics, 37*, 291–304.

Andreoni, J., & Vesterlund, L. (Accepted for publication). Public goods: A survey of experimental research. In J. Kagel, A. Roth (Eds.), *The Handbook of experimental economics*.

Becker, G. (1969). Crime and punishment: An economic approach. *Journal of Political Economy, 76*, 169–217.

Brandts, J., Cooper, D. J., & Fatas, E. (2007). Leadership and overcoming coordination failure whit asymmetric cost. *Experimental Economics, 10*(3), 269–284.

Brandts, J., Fatas, E., & Lagos, F. (2009). *An experimental study of equity and reciprocity using individual decisions*. Mimeo, CSIC-IAE.

Burgess, R., & Pande, R. (2005). Do rural banks matter? Evidence from the Indian social banking experiment. *American Economic Review, 95*(3), 780–795.

Che, Y. K., & Yoo, S. W. (2001). Optimal incentives for teams. *American Economic Review, 91*(3), 525–541.

Croson, R. (1996). Partners and strangers revisited. *Economics Letters, 53*, 25–32.

Croson, R. (2000). Feedback in voluntary contribution mechanisms: An experiment in team production. *Research in Experimental Economics, 8*, 85–97.

Croson, R., Fatas, E., & Neugebauer, T. (2005). Reciprocity, matching and conditional cooperation in two public goods games. *Economics Letters, 87*, 95–101.

Croson, R., Fatas, E., & Neugebauer, T. (2006). *Excludability and cooperation in team production*. Mimeo, LINEEX.

Fatas, E., Melendez-Jimenez, M. A., & Solaz, H. (2009). *Cooperation, punishment and networks*. Mimeo, LINEEX.

Fatas, E., & Godoy, S. (2009). *Take the lead!*. Mimeo, LINEEX.

Fatas, E., Godoy, S., & Solaz, H. (2009). *People with (out) history. Traceability and exemplar behavior*. Mimeo, LINEEX.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980–994.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71*, 397–404.

Fischbacher, U., & Gächter, S. (forthcoming). Social preferences, beliefs and the dynamics of free riding in public good experiments. *American Economics Review*.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.

Frechette, G. (2010). Learning in a multilateral bargaining game. *Journal of Econometrics, 153*(2), 183–195.

Frechette, G., Kagel, J. H., & Morelli, M. (2005). Behavioral identification in coalitional bargaining: An experimental analysis of demand bargaining and alternating offers. *Econometrica, 73*(6), 1893–1937.

Hamilton, B. H., Nickerson, J. A., & Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of team on productivity and participation. *Journal of Political Economy, 111*(3), 465–497.

Hermman, B., Gächter, S., & Thöni, C. (2008). Antisocial punishment across societies. *Science, 319*, 1362–1367.

Hirshleifer, D., & Rasmusen, E. (1989). Cooperation in a repeated prisoners' dilemma with ostracism. *Journal of Economic Behavioral and Organization, 12*, 87–106.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics, 13*, 324–340.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentives effects. *American Economic Review, 92*(5), 1655–1664.

Houser, D., Xiao, E., McCabe, K., & Smith, V. (2008). When punishment fails: Research on sanctions, intentions and non-cooperation. *Games and Economic Behavior, 62*(2), 509–532.

Hultsch, F. (1889). *The histories of Polybius*. London, New York: McMillan and Co..

Lazear, E. P. (1998). *Personnel economics for managers*. New York: John Wiley.

Ledyard, J. (1995). Public goods: A survey of experimental research. In J. Kagel & A. Roth (Eds.), *The handbook of experimental economics* (pp. 111–194). Princeton, NJ: Princeton University Press.

Margolis, H. (2007). *Cognition and extended rational choice*. Routledge.

McAfee, R. P., & McMillan, J. (1991). Optimal contracts for teams. *International Economic Review, 32*(3), 561–577.

Meran, G., & Schwalbe, U. (1987). Pollution control and collective punishments. *Journal of Institutional and Theoretical Economics, 143*(4), 616–629.

Miceli, T. J., & Segerson, K. (2007). Punishing the innocent along with the guilty: The economics of individual versus group punishment. *The Journal of Legal Studies, 36*(1), 81–106.

Nalbantian, H., & Schotter, A. (1997). Productivity under group incentives: An experimental study. *American Economic Review, 87*(3), 314–341.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics, 92*, 91–112.

Nikiforakis, N., & Normann, H. T. (2008). A comparative statics analysis of punishment in public good games. *Experimental Economics, 11*, 358–369.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *The American Political Science Review, 86*(2), 404–417.

Polinsky, A. M., & Shavel, S. (2000). The economic theory of public enforcement of law. *Journal of Economic Literature, 38*, 45–76.

Posner, R., & Rasmusen, E. (1999). Creating and enforcing norms, with special reference to sanctions. *International Review of Law and Economics, 19*, 369–382.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature, 7*, 63.

Rasmusen, E. (1987). Moral hazard in risk-averse teams. *Rand Journal of Economics, 18*(3), 428–435.

Sefton, M., Shupp, R., & Walker, J. (2005). *The effect of rewards and sanctions in provision of public goods*. Ball State University, Department of Economics, *Working Paper 200504*.

Shapiro, C., & Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review, 74*(3), 433–444.

Segerson, K. (1988). Uncertainty and incentives for nonpoint pollution control. *Journal of Environmental Economics and Management, 15*, 87–98.

Shortle, J. S., & Horan, R. D. (2001). The economics of nonpoint pollution control. *Journal of Economic Surveys, 15*(3), 251–253.

Smith, V. L. (1976). Experimental economics: induced value theory. *American Economic Review, 66*(2), 274–279.

Weber, R. A. (2006). Managing growth to achieve efficient coordination in large groups. *American Economic Review, 96*, 114–126.

Woolridge, J. (2003). Cluster-sample methods in applied econometrics. *American Economic Review, 93*, 133–138.

Xepapadeas, A. P. (1991). Environmental policy under imperfect information: incentives and moral hazard. *Journal of Environmental Economics and Management, 20*, 113–126.

Yamagishi, T. (1988). Seriousness of social dilemmas and the provision of a sanctioning system. *Social Psychology Quarterly, 51*(1), 32–42.