Exp Econ (2008) 11: 358–369
DOI 10.1007/s10683-007-9171-3

A comparative statics analysis of punishment in public-good experiments

Nikos Nikiforakis · Hans-Theo Normann

Received: 30 November 2005 / Revised: 12 April 2007 /
Accepted: 17 April 2007 / Published online: 29 September 2007

Abstract This paper provides a comparative-statics analysis of punishment in public-good experiments. We vary the effectiveness of punishment, that is, the factor by which punishment reduces the punished player's income. The data show that contributions increase monotonically in punishment effectiveness. High effectiveness leads to near complete cooperation and welfare improvements. Below a certain threshold, however, punishment cannot prevent the decay of cooperation. In these cases, punishment opportunities reduce welfare. The results suggest that the experimenter's choice of the punishment effectiveness is of great importance for the experimental outcome.

Keywords Decentralized punishment · Punishment effectiveness · Public good · Welfare

JEL Classification C92 · D70 · H41

1 Introduction

Recently, experimental economists devoted a lot of attention to the impact punishment opportunities have on cooperation in public-good games.[1] Punishment is de-

[1] See, for example, Ostrom et al. (1992), Fehr and Gächter (2000, 2002), Masclet et al. (2003), Carpenter (2007a), Nikiforakis (2005), Noussair and Tucker (2005), Page et al. (2005), Sefton et al. (2005) Anderson

Electronic supplementary material The online version of this article (http://dx.doi.org/10.1007/s10683-007-9171-3) contains supplementary material, which is available to authorized users.

N. Nikiforakis (✉)
Department of Economics, The University of Melbourne, Melbourne, Victoria 3010, Australia
e-mail: n.nikiforakis@unimelb.edu.au

H.-T. Normann
Department of Economics, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK
e-mail: hans.normann@rhul.ac.uk

centralized, as it is not carried out by a central authority, and costly, as individuals purchase and assign punishment "points" that reduce the recipient's income. The results show that the existence of punishment opportunities can improve the level of cooperation significantly.

Despite the large number of experiments, little is known about the robustness of the punishment schemes and the requirements on them to lead to higher cooperation levels and welfare improvement. Carpenter (2007b, p. 536) emphasizes this point and concludes that "there has been very little comparative statics analysis" and that the existing literature resembles "a series of unconnected islands".

The aim of this paper is to provide such a comparative statics analysis of punishment in public-good games. We study the effectiveness of punishment, that is, the amount by which a punishment point reduces the recipient's income. In particular, we examine four different levels of punishment effectiveness in addition to a standard public-good game without punishment opportunities.

The study of punishment effectiveness seems important for three reasons. First, one main conclusion of the literature on decentralized enforcement of cooperation is that the existence of punishment opportunities can lead subjects to "very high or even *full cooperation* ... whereas the same subjects converge to *full defection* in the no-punishment condition" (Fehr and Gächter 2000, p. 993; emphasis in the original). It is important to know, therefore, whether any degree of punishment effectiveness can have this effect or whether a specific minimum degree is required to sustain cooperation. Second, punishments are costly for both punishers and victims. The existence of punishment opportunities can lead to welfare improvements only if the benefits from higher cooperation outweigh the punishment costs. Therefore, even if a higher effectiveness of punishment would imply better cooperation, it is not obvious whether high effectiveness is better than low effectiveness from a welfare perspective. Third, researchers tend to choose different degrees of punishment effectiveness for their experiments. Therefore, it is useful to understand to what extent differences in results can be explained by the utilization of different degrees of effectiveness.

2 The experiment

The experiment is based on a repeated linear public-good game with n players, $n \geq 2$. In every period, each participant is given an endowment y. Players then decide simultaneously and without communication how much of the endowment to contribute to a public account, c_i, where $0 \leq c_i \leq y$. The rest $(y - c_i)$ remains in the player's own account. In addition to the money that player i keeps, i receives a fixed percentage of the group's total contribution to the public account, α, where $0 < \alpha < 1 < n\alpha$. This implies that the earnings of player i in the same period are

$$\pi_i = y - c_i + \alpha \sum_{i=1}^{n} c_i. \tag{1}$$

and Putterman 2006, Bochet et al. (2006), Cinyabuguma et al. (2006), Gürerk et al. (2006), Carpenter (2007b), Denant-Boemont et al. (2007).

In the treatments with punishment opportunities, a second stage is added. After participants decide how much to contribute to the public account, they are informed about how much the other individuals in their group contributed. They can then, if they wish, purchase punishment points to reduce the income of one or more other participants. Punishment is costly for the punisher as every point reduces his income by 1 ECU (experimental currency unit). Let p_{ij} denote the number of punishment points that player i assigns to j (where $i, j=1, \ldots, n$; $j \neq i$), and e the reduction that one punishment point causes to its recipient. Player i's earnings at the end of the period are accordingly

$$\pi_i = y - c_i + \alpha \sum_{i=1}^{n} c_i - \sum_{j \neq i} p_{ij} - e \sum_{j \neq i} p_{ji}. \qquad (2)$$

Notice that the experiment utilizes the linear punishment technology employed in several recent papers (Fehr and Gächter 2002; Sefton et al. 2005; Anderson and Putterman 2006; Carpenter 2007b; Nikiforakis et al. 2007).[2] The maximum number of points a participant can distribute to others is equal to his earnings from the first stage, that is, $\sum_{j \neq i} p_{ij} \leq y - c_i + \alpha \sum_{i=1}^{n} c_i$. As in stage one, punishment decisions are made simultaneously and without communication.

In all treatments, it is common knowledge that $y = 20$, $n = 4$ and $\alpha = 0.4$. The treatment variable is e, the effectiveness of punishment. We have $e \in \{1, 2, 3, 4\}$ and the treatments are labelled "1", "2", "3" and "4" respectively. As a control treatment, we use the public-good game without punishment. We label this treatment "0".

All treatments last for $T=10$ periods, a fact known to subjects. In the treatments with punishment opportunities, each subject is given a one-off lump-sum payment of 25 ECU to pay for any losses the participant might incur in the duration of the experiment.

For the experiment we use fixed (or "partners") matching, that is, the group composition is the same in all periods. The reason for choosing fixed rather than random matching, where the group composition changes in every period, is that previous public-good experiments with punishment opportunities have shown that participants use punishment more frequently when groups remain unchanged. Therefore, we expect the effect of punishment on contribution rates to be more pronounced, as, for example, in the "partners" sessions of Fehr and Gächter (2000), and differences across different degrees of punishment effectiveness clearer under fixed matching.[3]

Information feedback is as follows. Once the participants have contributed in stage one, they are informed about their group's total contribution to the public account, the individual contributions of the group members, and their own earnings as in (1).

[2]In some studies, a punishment point reduces the recipient's income by a percentage, typically 10% (Fehr and Gächter 2000; Masclet et al. 2003; Denant-Boemont et al. 2007; Nikiforakis 2005; Noussair and Tucker 2005). The effectiveness of the punishment points is convex in the target's income in these studies which makes this technology somewhat less attractive for comparative statics analysis. For a discussion of the non-linear punishment technology, see Casari (2005).

[3]One important contribution of the public-good experiments with punishment under "perfect-strangers" matching protocol is to show that costly punishments occur even though punishers will not meet punished participants again. See, for example, Fehr and Gächter (2002) and Egas and Riedl (2005).

To prevent the formation of individual reputation, every player is randomly given a number between 1 and 4 at the beginning of each period to distinguish their actions from those of the others in that period. Such a mechanism ensures that, even though the group members remain the same, the participants cannot create a link between the actions of the other subjects across the periods. In the treatments where punishment is available, at the end of each period, participants are informed about the punishment points they received in total, the associated income reduction and their earnings as in (2).

The experiments were conducted in the experimental laboratory of Royal Holloway College (University of London) and University College London in January and February 2005. We ran a total of ten sessions with a total of 120 subjects (12 per session). For each treatment, we have six groups, giving us six statistically independent observations. None of the participants had participated previously in a public-good experiment. Sessions lasted approximately fifty minutes and the average payment was £11.10 or roughly $20.90. The exchange rate between the experimental currency units and the British pound was 1 ECU = £0.04. The experiments were conducted using zTree (Fischbacher 2007).

3 Hypotheses

In the subgame perfect Nash equilibrium, punishments and punishment effectiveness do not matter. Under the standard assumption that individuals maximize their own monetary payoff, punishments will not occur in the last period of the experiment. Given that punishments do not take place, players do not contribute to the public account in the last period as this would only reduce their earnings (see (1)). By backward induction, this is the prediction for all periods. As punishments do not occur, the effectiveness of punishment has no effect. Similarly, contributions are predicted to be zero in treatment “0”. Therefore, in the subgame perfect Nash equilibrium, predictions are the same across treatments.

In addition to the unique subgame perfect Nash equilibrium of the finitely repeated game, there are imperfect Nash equilibria in which cooperation can occur and in which punishment and punishment effectiveness may play a role. Ostrom et al. (1992) derive such imperfect Nash equilibria for a common pool game with punishment possibilities. They propose a simple symmetric strategy which we modify here for our game. In every period except the final period, T, contribute $c_i = y$ and choose $p_{ij} = 0,\ j \neq i$. In the event of a deviation, play $c_i = 0$ and punish all players with $p_{ij} = \widetilde{p} > 0,\ j \neq i$, for one period, then resume to $c_i = y$ and $p_{ij} = 0$. In period T, play the static Nash equilibrium unless a deviation occurs in $T - 1$. This strategy is a Nash equilibrium as long as the punishment following a deviation is sufficiently severe. Whereas this argument shows that Nash equilibria with positive contribution rates exist, clearly many other equilibria exist, including one where no player contributes. Moreover, Nash equilibria with positive contributions are imperfect because they are based on non-credible threats. Nevertheless, the Ostrom et al. (1992) argument may suggest some positive link between punishment effectiveness and contributions. Intuitively, the higher e, the harsher the maximum

threat of punishment. Hence, there may be parameter constellations where imperfect Nash equilibria with positive contributions exist for some $\hat{e}$ but fail to exist for any $e < \hat{e}$. More specifically for our game, since the punishments points $\widetilde{p}$ assigned after a deviation must not exceed the minimum stage one income (that is, $\widetilde{p}(n-1) \leq \alpha y$), the harshness of punishments depends primarily on e. Therefore, for low e, the threat may not be sufficiently high to sustain cooperation in an imperfect Nash equilibrium.

Punishments and punishment effectiveness may also have an effect in the model of other-regarding preferences developed by Fehr and Schmidt (1999). In their model, people receive utility from their individual material income as in the standard model but their utility might be reduced from inequitable distributions of income. Fehr and Schmidt (1999, section IV) discuss predictions for a public-good game with punishment opportunities and include predictions about the relation between the effectiveness of punishment and contribution behavior.[4] In particular, Fehr and Schmidt (1999) show that, as e increases, the requirement for the existence of equilibria with positive contributions is relaxed. Provided this requirement is met, actually any level between zero contributions and contributions of the full endowment can be sustained in a subgame perfect equilibrium. However, if the condition is not met, the only subgame perfect equilibrium remaining is the one described above with zero contributions. It follows that higher levels of punishment effectiveness make more likely the existence of equilibria with positive contributions.

4 Results

We start by reporting results on contributions to the public account, followed by a welfare comparison of the treatments, and a brief discussion of punishment behavior. Unless otherwise noted, we use data from each group as one observation. The group data are summarized in Table 1 in the Appendix. Since we have more than two treatments, we report Kruskal-Wallis tests when we compare means. We proceed to pair-wise treatment comparisons using the Dwass-Steel-Critchlow-Fligner *post-hoc* procedure which corrects for multiple comparisons (see e.g. Hollander and Wolfe 1999) only if the Kruskal-Wallis indicates significant differences in the means.

Figure 1 presents the mean contributions in periods 6–10 across treatments. Figure 1 reveals that, as the effectiveness of punishment increases, so does the mean contribution. The relationship between effectiveness and contributions is monotonic; we never find that higher effectiveness leads to lower contributions. Average treatment contribution rates and treatment effectiveness are positively correlated (Spearman correlation, two sided, $p = 0.01$). The differences in mean contributions across the treatments are also statistically significant (Kruskal-Wallis, $d.f. = 4$, $p = 0.01$).[5] We summarize

[4] Fehr and Schmidt (1999, section IV) analysis of public-good games with punishment is actually formulated in terms of the cost of punishment, not punishment effectiveness. The cost of punishment is the amount of ECU that a player must pay to reduce one's income by 1 ECU—in our case $1/e$. As pointed out by a referee, while cost of punishment and punishment effectiveness are analytically equivalent, this may not be the case behaviorally.

[5] The Dwass-Steel-Critchlow-Fligner *post-hoc* procedure indicates that five out of ten pair-wise comparisons confirm the monotonic relationship between effectiveness and contributions and none contradicts it.

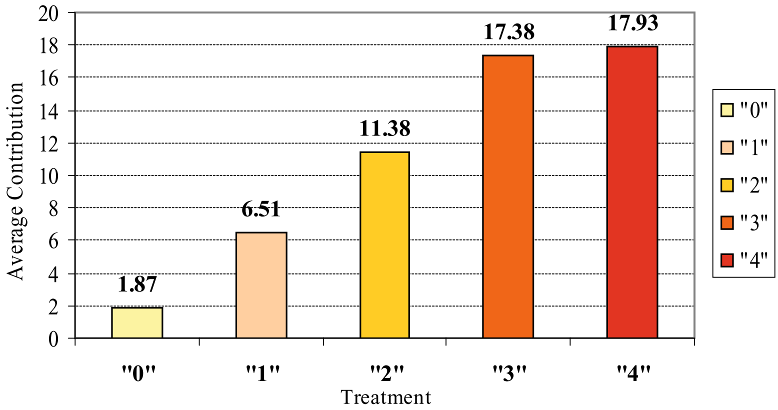

Fig. 1 Average contribution by treatment (data from periods 6–10)

Result 1 *Average contributions increase monotonically in the effectiveness of punishment.*

The evolution of average contributions over time is illustrated in Fig. 2. In period one, subjects contribute on average between 40% and 63% of their endowment, consistent with previous public-good experiments with and without punishment. Period one contributions are not significantly different across treatments (Kruskal-Wallis, $d.f. = 4$, $p = 0.610$). From period two onwards, contributions decrease in treatments "0" and "1", are roughly constant in "2", whereas they increase in treatments "3" and "4". The time trends are significant except for treatment "2" (Spearman correlation, two tailed, $p < 0.05$).[6] Remarkably, Result 1 (monotonicity) holds in every period except for the first.

Result 2 *Average contributions decrease over time in "0" and "1", they are constant in "2", whereas they increase in "3" and "4".*

We now discuss the effect of punishment effectiveness on subjects' earnings. Higher contributions do not necessarily imply higher earnings in experiments with punishment. The reason is that punishments are costly for both punishers and punished individuals. The welfare loss due to punishments might offset the welfare gain from higher contributions. To see how the effectiveness of punishment influences group welfare (which is equal to the sum of earnings), we calculate the cumulative

Treatments "2", "3" and "4" have significantly higher contribution rates than "0" (all $p = 0.05$, two-sided) and treatments "3" and "4" have significantly higher contribution rates than "1" ($p = 0.1$, two-sided). Note that this procedure is more conservative than the Mann-Whitney procedure which is often used for pairwise comparisons across treatments.

[6]The Spearman correlations between period and average contribution are $\rho = -0.988$, $p < 0.001$ ("0"); $\rho = -0.842$, $p = 0.002$ ("1"); $\rho = -0.489$, $p = 0.248$ ("2"); $\rho = 0.730$, $p = 0.017$ ("3"); $\rho = 0.679$, $p = 0.025$ ("4").

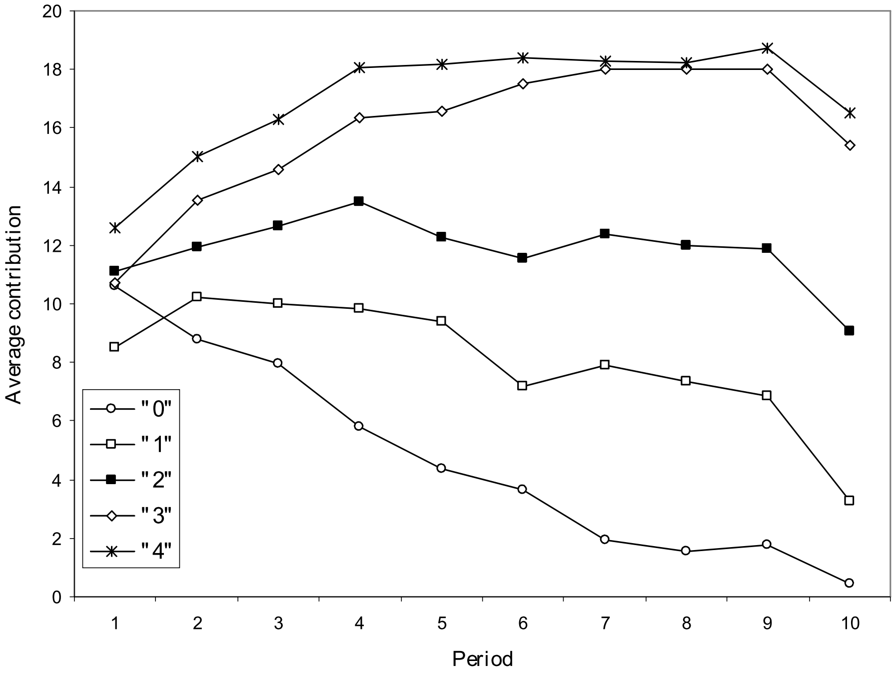

Fig. 2 Average contribution over time across treatments

relative earnings (*CRE*) between treatments with punishment and treatment "0". This is done using the following formula

$$CRE_{\tau}^{t} = \left(\sum_{s=1}^{t} \bar{\pi}_{\tau}^{s} - \sum_{s=1}^{t} \bar{\pi}_{0}^{s}\right) \Big/ \sum_{s=1}^{t} \bar{\pi}_{0}^{s} \tag{3}$$

where CRE_{τ}^{t} stands for cumulative relative earnings in treatment $\tau \in \{$"1", "2", "3", "4"$\}$ compared to treatment "0" up to period t, and $\sum_{s=1}^{t} \bar{\pi}_{\tau}^{s}$ the average cumulative earnings up to period t of treatment τ.

The evolution of *CRE* over time can be seen in Fig. 3. In period one, the treatments with punishment opportunities have $CRE < 0$ as contribution rates are similar to those in "0" and individuals incur the punishment-associated costs. From period one onwards, *CRE* increases in all treatments. However, as Fig. 3 illustrates, the increase in "1" is not sufficient to offset the losses that occurred in the first periods. In "2", higher contributions manage to offset the costs of punishment only in the last two periods. This holds true for treatment "3" from period six onwards. Finally, we see treatment "4" has the highest level of welfare as $CRE > 0$ already in period four. The difference in total earnings across all ten periods, as observed in $t = 10$ in Fig. 3, is statistically significant (Kruskal-Wallis, $d.f. = 4$, $p = 0.02$).[7]

[7]The Dwass-Steel-Critchlow-Fligner *post-hoc* procedure indicates that treatments "3" and "4" have significantly higher cumulative group payoffs in period ten than "0", and that treatments "4" has significantly higher cumulative group payoffs than "1" (all $p = 0.05$, two-sided).

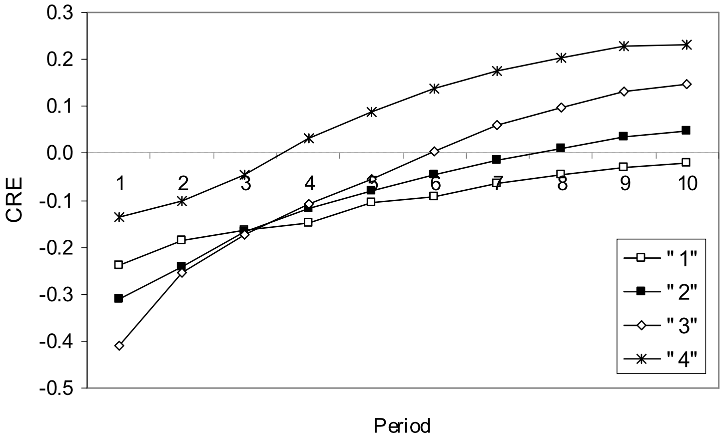

Fig. 3 Cumulative relative earnings over time across treatments

Result 3 *A punishment effectiveness of "3" or greater is required to obtain a welfare improvement compared to the public-good game without punishment* (*treatment "0"*). *The higher punishment effectiveness is*, *the sooner the existence of punishment opportunities leads to a welfare improvement compared to treatment "0"*.

Fehr and Gächter (2000, p. 993) make a point which may qualify Result 3. They argue that behavior in the second half of the experiments may represent the steady state. For our data, Table 1 in the Appendix shows that earnings in all treatments with punishment opportunities are higher than "0" in periods 6 to 10. Thus, according to the Fehr and Gächter (2000) argument, all our punishment treatments would yield higher welfare than in "0" with a horizon sufficiently longer than $T = 10$. However, this need not be the case. Page et al. (2005) use a public-good game with punishment for twenty periods. They find that welfare is still lower in the punishment treatment. Further, contributions in public-goods experiments without punishment may not decay as quickly when there are many periods (Keser and van Winden 2000). Hence, whether punishment opportunities yield welfare improvements irrespective of punishment effectiveness in the long run remains an open question.

We also analyzed the determinants of punishment in detail, but we refrain from reproducing the regression results here as we, by and large, confirm previous results (Ostrom et al. 1992; Anderson and Putterman 2006; Carpenter 2007b). For a detailed discussion of the findings see Nikiforakis and Normann (2005). In summary, we find a positive relationship between the effectiveness of punishment and the punishment that i inflicts on j. Given that the cost of punishment is $1/e$, this verifies previous characterizations of punishment as an ordinary good.[8] Also in line with the existing literature, we find that the higher the income of i, the less punishment points i

[8]One of our referees is sceptical about interpreting punishment as a good for the following reasons. On top of the question whether the cost of punishment can actually be viewed as a price, the referee correctly points out that the existence of a downward sloping demand curve at the aggregate level does not imply that the law of demand holds at the individual level. Moreover, the demand interpretation is based on the rational choice model, however, individuals are not always fully rational in these experiments.

assigns. That is, punishment can be characterized as an inferior good (as defined in Varian 1999). Across treatments, individuals are more severely punished the lower their contribution is compared to that of their peers, while punishments decline with time. Finally, we observe considerable amounts of punishment in treatment "1" although punishments in that treatment can not reduce income differences between participants (a point which has also been made in Falk et al. 2005; Sefton et al. 2005; Anderson and Putterman 2006; Masclet and Villeval 2006). This behavior indicates that punishments cannot solely be explained by the model of inequity aversion of Fehr and Schmidt (1999).[9]

5 Discussion and conclusion

This paper shows that the selection of punishment effectiveness (defined as the factor by which one punishment point reduces the punished player's income) plays a significant role in determining the experimental outcome. Contributions to the public account increase monotonically in the effectiveness of punishment as does welfare. A minimum level of punishment effectiveness (three) is required to significantly raise contributions over time. Similarly, a punishment effectiveness of two is required so that the benefits of higher contribution rates offset the costs of punishment. Note that we do not claim that these statements hold regardless of other aspects of the experimental design. For example, Bochet et al. (2006) and Page et al. (2005) employ a punishment effectiveness of four and find moderate cooperation levels. Whatever may account for the different results (there are numerous differences in the design and the procedures), the point of our study is to show that cooperation and welfare unambiguously increase with the effectiveness of punishment, *ceteris paribus.*

Independently of us, Egas and Riedl (2005) have produced a related study. Using a public-good game with punishment opportunities under a "perfect-strangers" matching protocol, they examine the effect of two variables: The income reduction caused by punishment and the cost of purchasing one punishment point. Egas and Riedl's experimental design includes treatments with punishment effectiveness $1/3$, 3, and two treatments with a punishment effectiveness of 1.[10] Their results confirm ours by showing that as punishment effectiveness increases so does cooperation, even though the difference between treatments with effectiveness of $1/3$ and 1 is not significant. The impact of effectiveness on welfare has the opposite sign compared to our results. This is likely to be due to the low cooperation levels under perfect-strangers.

[9]Individual behavior resembles the strategy of the imperfect Nash equilibria in that punishments usually follow instances of low contributions. However, similar to Ostrom et al. (1992, p. 411), we also find inconsistencies with the imperfect Nash equilibria in that punishments were too severe and earnings too low in many cases.

[10]In our experiment, we keep the nominal price of buying a punishment point fixed at one. By altering the income reduction caused by a single punishment point we get different levels of effectiveness. Egas and Riedl (2005) have a treatment where a punishment point costs one experimental currency unit to the punisher and reduces the income of the punished player also by one unit. Then they have a treatment where a punishment point costs three units to the punisher and reduces the income of the punished also by three units. Both treatments, in our setup, imply an effectiveness of one. Egas and Riedl (2005) do not find any significant differences between the two treatments.

Two further related papers are Anderson and Putterman (2006) and Carpenter (2007b). These papers vary the cost of punishment, but, in contrast to our approach, they mainly vary the cost of punishment *within* treatments whereas we vary punishment of effectiveness *across* treatments. In our experiment, participants make their decisions unaware of alternative degrees of punishment effectiveness. Nevertheless, the results from Anderson and Putterman (2006) and Carpenter (2007b) are consistent with our findings.

Acknowledgements The authors would like to thank Jeff Carpenter, Dirk Engelmann, Ernst Fehr, Simon Gächter, Steffen Huck, an anonymous referee, and the participants in the European Meeting of ESA in Alessandria in 2005 for their helpful comments. The authors also gratefully acknowledge support from ESRC (project RES-000-22-0948).

Appendix

Table 1 Group data

Treatment	Group	Mean contribution		Mean earnings	
		all periods	periods 6–10	all periods	periods 6–10
0	1	3.70	0.35	22.22	20.21
0	2	2.95	0.50	21.77	20.30
0	3	7.85	3.55	24.71	22.13
0	4	4.38	3.05	22.63	21.83
0	5	2.53	1.55	21.52	20.93
0	6	6.73	2.20	24.04	21.32
0	average	4.69	1.87	22.81	21.12
1	11	14.45	14.00	24.72	25.10
1	12	9.20	4.50	19.92	21.50
1	13	3.08	2.30	20.90	20.48
1	14	15.63	15.50	25.63	27.80
1	15	3.88	1.90	22.13	21.04
1	16	2.05	0.85	20.68	20.51
1	average	8.05	6.51	22.33	22.74
2	21	18.08	19.00	28.15	29.60
2	22	10.53	11.45	23.99	25.97
2	23	10.53	9.15	19.12	20.84
2	24	7.23	4.90	21.34	21.14
2	25	19.33	20.00	29.05	32.00
2	26	5.33	3.75	21.77	21.95
2	average	11.83	11.38	23.90	25.25
3	31	16.00	18.00	24.60	28.40
3	32	16.75	18.10	27.95	30.46
3	33	13.73	17.50	22.34	28.10

Table 1 (*Continued*)

Treatment	Group	Mean contribution		Mean earnings	
		all periods	periods 6–10	all periods	periods 6–10
3	34	19.25	20.00	28.85	32.00
3	35	17.90	19.00	26.64	29.20
3	36	11.58	11.70	26.55	26.62
3	average	15.87	17.38	26.15	29.13
4	41	10.48	12.38	25.54	27.11
4	42	19.40	20.00	29.77	32.00
4	43	18.85	19.25	29.69	29.68
4	44	18.58	18.75	29.52	28.75
4	45	17.63	18.44	26.83	28.25
4	46	17.23	18.75	27.21	28.75
4	average	17.03	17.93	28.09	29.09

References

Anderson, C., & Putterman, L. (2006). Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior*, *54*(1), 1–24

Bochet, O., Page, T., & Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior and Organization*, *60*(1), 11–26

Carpenter, J. (2007a, in press). Punishing free-riders: how group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*

Carpenter, J. (2007b). The demand for punishment. *Journal of Economic Behavior and Organization*, *62*, 522–542

Casari, M. (2005). On the design of peer punishment experiments. *Experimental Economics*, *8*(2), 107–115

Cinyabuguma, M., Page, T., & Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics*, *9*, 265–279

Denant-Boemont, L., Masclet, D., & Noussair, C. (2007, in press). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory*

Egas, M., & Riedl, A. (2005). *The economics of altruistic punishment and the demise of cooperation*. Tinbergen institute discussion Papers 05-065/1

Falk, A., Fehr, E., & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica*, *73*(6), 2017–2030

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

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition and co-operation. *Quarterly Journal of Economics*, *114*, 817–868

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

Gürerk, O., Irlendbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, *312*, 108–111

Hollander, M., & Wolfe, D. A. (1999). *Nonparametric statistical methods*. New York: Wiley

Keser, C., & van Winden, F. (2000). Partners contribute more than strangers: conditional cooperation. *Scandinavian Journal of Economics*, *102*(1), 23–39

Masclet, D., Noussair, C., Tucker, S., & Villeval, M. C. (2003). Monetary and non-monetary punishment in the voluntary contributions mechanism. *American Economic Review*, *93*, 366–380

Masclet, D., & Villeval, M. C. (2006). *Punishment, inequality and emotions*. IZA Discussion Paper No. 2119

Nikiforakis, N. (2007, in press). Punishment and counter-punishment in public good games—can we really govern ourselves? *Journal of Public Economics*

Nikiforakis, N., & Normann, H. T. (2005). *A comparative statics analysis of punishment in public-good experiments*. Royal Holloway Discussion Papers Series, dpe05/7

Nikiforakis, N., Normann, H. T., & Wallace, B. (2007). *Asymmetric enforcement of cooperation in a social dilemma*. University of Melbourne Economics Working Paper No. 982

Noussair, C., & Tucker, S. (2005). Combining monetary and social sanctions to promote cooperation. *Economic Inquiry*, *43*(3), 649–660

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

Page, T., Putterman, L., & Unel, B. (2005). Voluntary association in public goods experiments: reciprocity, mimicry and efficiency. *The Economic Journal*, *115*(506), 1032–1053

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

Varian, H. R. (1999). *Intermediate microeconomics* (5th ed.). New York: Norton