

Peer punishment across payoff equivalent public good and common pool resource experiments

David C. Kingsley¹

Received: 7 May 2015 / Revised: 29 October 2015 / Accepted: 3 November 2015 /
Published online: 19 November 2015
© Economic Science Association 2015

Abstract This paper investigates the effectiveness of peer punishment in non-linear social dilemmas and replicates Cason and Gangadharan (Exp Econ 18:66–88, 2015). The contribution of this replication is that cooperation is quantified across payoff equivalent, strategically symmetric public good and common pool resource experiments. Results suggest that the cooperation-inducing effect of peer punishment is statistically equivalent across conditions. Despite this increase in cooperation, earnings are significantly lower than in the absence of punishment. Institutional features which improve the effectiveness of peer punishment in linear public good experiments may, similarly, make self-governance possible in more complex social dilemmas.

Keywords Public good · Common pool resource · Experiment · Cooperation · Self-governance

JEL Classification C92 · D70 · H41

Electronic supplementary material The online version of this article (doi:[10.1007/s40881-015-0017-z](https://doi.org/10.1007/s40881-015-0017-z)) contains supplementary material, which is available to authorized users.

✉ David C. Kingsley
david_kingsley@uml.edu

¹ Department of Economics, University of Massachusetts Lowell, 1 University Ave., Lowell, MA 01854, USA

1 Introduction

Much of the research suggesting that groups can self-govern comes from linear, boundary solution, public good experiments (Chaudhuri 2011). Such experiments are useful because they are easily explained to subjects and provide the starkest possible equilibrium predictions. That is, the incentives are all or nothing; self-interest suggests complete free-riding while efficiency requires contributing one's entire endowment. However, social dilemmas observed in the real world are rarely as stark as those presented in linear public good experiments. It is thus important to understand whether peer to peer punishment is as effective in non-linear, i.e. more complex, environments.

Cason and Gangadharan (2015) investigated the effectiveness of peer to peer punishment in non-linear public good (PG) and common pool resource (CPR) experiments. In both treatments, the opportunity to punish increased cooperation. However, the positive impact was too weak to increase welfare (group earnings). Cason and Gangadharan suggest that the non-linear environment limits the effectiveness of peer punishment because the social optimum is harder to identify and, therefore, to coalesce around. These results suggest that cooperative behavior is qualitatively similar across public good and common pool resource experiments. However, the treatments investigated by Cason and Gangadharan are not payoff equivalent and thus it is difficult to compare behavior across treatments.

This paper replicates the findings of Cason and Gangadharan (2015) across payoff equivalent, strategically symmetric, PG and CPR experiments. The contribution of this replication is that the level of cooperation can be quantified and directly compared across conditions. Results suggest that peer punishment induces the equivalent amount of cooperation in both treatments. Despite the increased cooperation, punishment alone does not improve welfare. However, the equivalence of the cooperation-inducing effect across treatments suggests that the same procedures that increase the effectiveness of peer punishment in linear PG experiments may have a similar impact within more complicated social dilemmas. This suggests that self-governance in complex social dilemmas is possible.

2 The experiment

2.1 Experimental design

This paper investigates behavior across PG and CPR experiments with and without the opportunity to punish. Treatments are referred to as PGnoPun, CPRnoPun, PGPun, and CPRPun. All subjects were recruited from the University of Massachusetts Amherst. In each session, 4 groups of 4 students were randomly formed and maintained throughout. At the beginning of each period, each subject was given an endowment of 50 Experimental Dollars (EDs) which they were able to allocate between two accounts. All subjects were instructed to allocate any integer

of their EDs to account 1 (the group account) and that the remainder of their endowment would be allocated to account 2 (their private account).

Instructions displayed individual and group payoff tables describing how individual and group period earnings varied across the range of group account allocations.¹ Individual and group payoff tables were included to ensure that each subject understood the relationship between their individual allocation, their individual earnings, and the earnings of the group. Subjects played one practice period to demonstrate how the game would proceed for the next 15 paid periods. In all treatments, subjects were required to allocate 25 EDs to the group account in the practice period. As shown below, in all treatments, a group account allocation of 25 EDs from each group member maximized and equally distributed group earnings.

After each period, subjects were shown their allocation to the group account, the total allocation to the group account, their period earnings, and their accumulated earnings. They were also shown the allocations of each group member in random order. The order of the presentation of individual allocations was randomized each period to mitigate reputation effects. They were paid privately at the end of the experiment such that 100 ED = \$1.

The payoff function in the PG treatments is:

$$\pi_i = w(e - x_i) + \frac{1}{n} \left[a \sum_{j=1}^n x_j - b \left(\sum_{j=1}^n x_j \right)^2 \right] \quad (1)$$

where $w = 1$ is the return from the private account, $e = 50$ is the subject's endowment, $a = 6$, $b = 0.025$, $n = 4$ and x_i is the subject's group account allocation. The second term represents each subject's share of the return from the group account. This payoff function indicates that it is in the self-interest of each subject to allocate EDs to the group account until $\sum x_j = \frac{a-w}{2b} = 40$. Thus, the symmetric Nash allocation is 10 EDs.

The payoff function in the CPR treatments is:

$$\pi_i = w(e - x_i) + \frac{x_i}{\sum_{j=1}^n x_j} \left[a \sum_{j=1}^n x_j - b \left(\sum_{j=1}^n x_j \right)^2 \right] \quad (2)$$

where, compared to the payoff function in the PG treatments, the only difference is how the returns from the group account are distributed.² Again, the second term represents each subject's share of the return from the group account which is now proportional to their allocation. This payoff function indicates that it is in the self-interest of each subject to allocate EDs to the group account until $\sum x_j = \left(\frac{n}{n+1} \right) \frac{a-w}{b} = 160$. Thus, the symmetric Nash allocation is 40 EDs.

¹ Complete experimental instructions can be found in the Online Appendix.

² This CPR payoff structure was first introduced by Ostrom et al. (1992) and suggests that each subject has an incentive to over allocate to the group account. That is, each subject will over-use the resource as they ignore the cost that their allocation imposes on others.

Table 1 Experimental design

	Public good	Common pool resource
Endowment	50	50
Private return	1	1
Group return	$(1/n)F(\sum x_j)$	$(x_i/\sum x_j)F(\sum x_j)$
Sym. nash allocation	10	40
Sym. nash payoff	90	90
Sym. social allocation	25	25
Sym. social payoff	112.5	112.5

In all treatments, randomly formed groups of 4 subjects played for 15 periods. $F(\sum x_j) = 6 \sum x_j - 0.025(\sum x_j)^2$. Values are expressed in Experimental Dollars

The social optimum is determined by treating the group as a single entity such that the group account return accrues to the group rather than being distributed across group members. From this perspective, it is in the interest of the group to equate the return from the private account and the aggregate return from the group account. At the group level, the unique social optimum allocation is such that $\sum x_j = \frac{a-w}{2b} = 100$ EDs are allocated. Thus, in both the PG and CPR treatments, the symmetric social optimum allocation is 25 EDs. In all treatments, participants earn 112.5 EDs at the symmetric social optimum each period. The symmetric Nash allocations are equally distant from this social optimum at 10 and 40 EDs in the PG and CPR treatments, respectively. In all treatments, participants earn 90 EDs at the symmetric Nash equilibrium each period. Table 1 summarizes the experimental design.³

Subjects in PGPun and CPRPun were given the same instructions as their no-punishment counterparts except that their game was augmented with a second decision stage. After each group member made his/her allocation decision, each subject continued to be shown the individual allocations of each group member in random order. Each subject now had the opportunity to pay 1 ED to punish another group member by reducing his/her period earnings by 3 EDs. They were allowed to place multiple sanctions on any number of other group members but had to pay the cost from the current period's earnings.

³ In Cason and Gangadharan (2015), groups of 4 subjects played for 10 periods and had an endowment of 12 tokens in each treatment. In their non-linear PG, the return from the group account was constant at 1 per token allocated. The private account return was 5 for the first 2, 2 for the next seven (3–9), and 0.8 for the last three (10–12) tokens allocated. Thus, the Nash equilibrium allocation to the group account was 3 and resulted in a 36 token payoff. The social optimum allocation to the group account was 10 and resulted in a 50 token payoff. In their non-linear CPR the return from the private account was constant at 2 per token. The return from the group account was similar to the established literature: $(x_i/\sum x_j)(18 \sum x_j - 0.4(\sum x_j)^2)$. Here, the Nash and social optimum group account allocations (payoffs) were 8 and 5 (37.6 and 52) tokens, respectively. The punishment mechanism cost 1 token and imposed a sanction of 3 tokens. Each subject could place up to 5 sanctions per subject per period.

Table 2 Demographics and average allocations, cooperation scores, and earnings

	Male	GPA	N	Allocation	Cooperation score	Earnings
<i>No punishment</i>						
Public good	0.047	3.45	32	17.2 (1.14)	34.35 (2.28)	\$20.64 (0.36)
Common pool resource	0.56	3.18	32	33.2 (2.48)	33.61 (4.97)	\$20.54 (0.66)
<i>Punishment</i>						
public good	0.63	3.36	32	20.8 (2.72)	41.54 (5.44)	\$21.19 [\$19.01] (0.36) [(1.61)]
Common pool resource	0.59	3.32	32	29.5 (3.04)	41.01 (6.08)	\$21.08 [\$17.25] (0.49) [(1.54)]

Earnings, after the cost of punishment, are shown in brackets. Standard deviations, calculated using the group average as the unit of observation, reported in parentheses. The cooperation score is calculated as $(x_i/50) \times 100$ in the PG conditions and $(1 - x_i/50) \times 100$ in the CPR conditions

2.2 Replication study power calculations

The current study first tests whether the findings presented in Cason and Gangadharan (2015) are replicated. These findings are a significant increase in cooperation with than without punishment in both the PG and CPR games. Given our sample size of 8 independent observations per group, and the observed means and standard deviation in the original study, we have the following power:⁴

- cooperation is significantly higher at the $p < 5\%$ level 14.98 % of the time with than without punishment in the PG game,
- cooperation is significantly higher at the $p < 5\%$ level 34.75 % of the time with than without punishment in the CPR game.

3 Results

A total of 128 subjects participated in the experiment providing a total of 1920 observations. Table 2 displays the collected demographic variables [self-reported grade point average (GPA) and Gender (male = 1)] and the average earnings (after the costs of punishment in brackets) in each treatment. In PGnoPun and CPRnoPun, average earnings were \$20.64 and \$20.54, respectively. In PGPun and CPRPun, average earnings were \$21.19 and \$21.08 before, and \$19.01 and \$17.25 after, the costs of punishment, respectively.

⁴ Cason and Gangadharan (2015) have 12 independent group level observations and observe the following. Average (standard deviation) group account allocations: PG 5.69 (1.53), PG Pun. 6.47 (1.87), CPR 7.29 (0.74), CPR Pun. 6.80 (0.49).

Table 3 Punishment behavior

	Punishment decision	Punishment level
Public good treatment	-0.043 (0.080)	-5.74 (4.96)
Below average cooperation	0.007*** (0.002)	0.583*** (0.124)
Above average cooperation	-0.001 (0.002)	0.181*** (0.056)
Period average cooperation	-0.003 (0.002)	0.002 (0.099)
Period	-0.012*** (0.004)	-0.505*** (0.218)
Constant	1.24*** (0.41)	2.27*** (0.29)
<i>N</i>	960	693
Wald χ^2	55.51***	104.09***

Hurdle model. The punishment decision is estimated using a Probit specification, the dependent variable equals 1 if the subject received punishment and zero otherwise. If the subject is punished, the punishment level is estimated using a truncated exponential regression. Average marginal effects are reported. Standard errors are clustered by group to account for within-group correlation and are reported in parentheses. Results are robust to including period indicators rather than as a linear trend. Coefficients significantly different from zero at 10 % (*), 5 % (**), and 1 % (***)

To quantify the level of cooperation on a common metric an index is developed which captures, on a 0–100 scale, the level of cooperation in each treatment (Table 2). In the PG treatments, the index is $(x_i/50) \times 100$ and in the CPR treatments, it is $(1 - x_i/50) \times 100$. In all treatments, the symmetric social allocation (25 EDs) yields a cooperation score of 50. The symmetric Nash allocations (10 and 40 EDs in the PG and CPR treatments, respectively) yield a cooperation score of 20. The importance of the index is to capture the level of cooperation across treatments such that each index score implies the same relative level of cooperation. Table 2 displays the average group account allocations and the average cooperation scores. Mann–Whitney *U* tests (MW) are presented to investigate average cooperation scores across conditions using group average as the unit of observation (8 independent observations per condition). In line with the qualitative analysis of Cason and Gangadharan (2015), average cooperation scores are predicted to be statistically equivalent with and without punishment.

As shown in Table 2, the average cooperation score was 34.35 and 33.61 in PGnoPun and CPRnoPun, respectively, and was not significantly different (MW $n = 16$; $z = 0.263$; $p = 0.793$).⁵ Similarly, average earnings are not significantly different across PGnoPun and CPRnoPun at \$20.64 and \$20.54, respectively (MW $n = 16$; $z = 0.105$; $p = 0.916$). Therefore, across payoff equivalent PG and CPR experiments cooperation and welfare are statistically equivalent in the absence of punishment.

The opportunity to punish significantly increases average allocations (cooperation scores) from 17.2 EDs to 20.8 EDs (from 34.35 to 41.54) in the PG treatments (MW $n = 16$; $z = 2.522$; $p = 0.012$) and significantly decreases (increases) average allocations (cooperation scores) from 33.2 EDs to 29.5 EDs (from 33.61 to 41.01) in

⁵ Note that given the sample size of 8 per treatment the difference required to detect it as being significant at the 5 % level with 80 % probability is 1.51.

the CPR treatments (MW $n = 16$; $z = 2.310$; $p = 0.021$). As a result, peer punishment induces statistically equivalent levels of cooperation in both treatments (MW $n = 16$; $z = 0.525$; $p = 0.599$). After the costs of punishment are included average earnings are significantly lowered from \$20.64 to \$19.01 in the PG treatments (MW $n = 16$; $z = 3.048$; $p < 0.01$) and from \$20.54 to \$17.25 in the CPR treatments (MW $n = 16$; $z = 2.521$; $p = 0.012$). Just as in the absence of punishment average earnings are statistically equivalent across the punishment treatments (MW $n = 16$; $z = 0.105$; $p = 0.916$).

Beyond the effect of punishment on cooperation, it is of interest to investigate the use of punishment across treatments. Following Nikiforakis (2008), a hurdle model is estimated to understand both the decision to punish, as a binary choice, and the amount of punishment received. The independent variables include an indicator for the PG treatment, the amount that subject j 's cooperation score deviates above and below the group's period average cooperation score, the period's average cooperation score, and the period to account for a time trend.

Column 1 (Table 3) displays the average marginal effects of the independent variables on the probability of receiving punishment. Results suggest that a below-average cooperation score increases the probability that one is punished and that the probability of being punished decreases with repetition. Importantly, the PG treatment indicator is insignificant suggesting that there is no systematic variation in one's probability of being punished across treatments.

Column 2 (Table 3) displays the average marginal effects of the independent variables on the amount of punishment received. Here, cooperation scores above and below the group's period average increase the amount of punishment received. While punishing group members who cooperate more than average is not intuitive note that the punishment received is 3.2 times higher when one's cooperation score is below average. For example, a one unit increase (decrease) above (below) the group's period average cooperation score increases the amount of punishment received by 0.18 (0.58) EDs. Further, when the model is restricted to only those observations for which the sum of the groups' cooperation scores is 2 or less the impact of above average cooperation scores is only marginally significant (0.18 (0.11)*).⁶ Period remains negative and significant suggesting that the level of punishment decreases with repetition.

4 Discussion and conclusion

This paper investigated the effectiveness of peer punishment across payoff equivalent PG and CPR experiments. Using an index to quantify the level of cooperation, results suggest that the cooperation-inducing effect of peer punishment is quantitatively equivalent across treatments. Despite the increased cooperation, the

⁶ In periods where the sum of the members cooperation scores are greater than 2 one or more group members are, in effect, being too cooperative. For example, in the PG treatment, this means that group account allocations are above 100 EDs (the social optimum level). Because contributions above the social optimum reduce the returns from the group account punishment of high contributors is in the subject's self-interest.

opportunity to punish reduces welfare in both treatments. Importantly, the similarity observed across treatments suggests that there is no fundamental difference in a groups' capacity for cooperation across PG and CPR experiments.

The linear PG literature suggests that peer punishment is more effective when (1) established groups interact over longer durations (Gächter et al. 2008), at least when counter-punishment is not possible (Engelmann and Nikiforakis 2015), in and (2) are able to employ a cost-effective, high impact, form of punishment (Sefton et al. 2007; Nikiforakis and Normann 2008; Egas and Riedl 2008). Whether the required duration, with or without counter-punishment, and impact of punishment are positively related to complexity, however measured, is an open empirical question. These results motivate future research to determine the capacity for self-governance in complex social dilemmas.

Acknowledgements I would like to thank Alexander Smith, Brock Stoddard, participants at the 2013 North American Economic Science Association conference, the editor, Nikos Nikiforakis, and two anonymous reviewers. Financial support provided by the University of Massachusetts Lowell is gratefully acknowledged. I would also like to thank John Spraggon and the staff at the Cleve E. Willis Experimental Laboratory at the University of Massachusetts Amherst.

References

- Cason, T., & Gangadharan, L. (2015). Promoting cooperation in nonlinear social dilemmas through peer punishment. *Experimental Economics*, 18, 66–88.
- Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14(1), 47–83.
- Egas, M., & Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 275, 871–878.
- Engelmann, D., & Nikiforakis, N. (2015). In the long run we are all dead: On the benefits of peer punishment in rich environments. *Social Choice and Welfare*, 45(3), 561–577.
- Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science*, 322, 1510.
- 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 experiments. *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, 404–417.
- Sefton, M., Shupp, R., & Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry*, 45(4), 671–690.