Journal of the Economic Science Association (2024) 10:603–608
https://doi.org/10.1007/s40881-023-00157-z

REPLICATION PAPER

Who punishes? A note on responses to cooperation and defection across cultures

Talbot M. Andrews[1]

Received: 12 February 2020 / Revised: 21 November 2023 / Accepted: 15 December 2023 /
Published online: 6 February 2024

Abstract

While people are surprisingly cooperative in social dilemmas, cooperation is fragile to the emergence of defection. Punishment is a key mechanism through which people sustain cooperation, but when are people willing to pay the costs to punish? Using data from existing work on punishment in public goods games conducted in industrialized countries throughout the world (Herrmann et al. in Science, 319(5868):1362–1367, 2008. https://doi.org/10.1126/science.1144237), I find first that those who contribute more are consistently punished less. Second, in many study locations, there are insignificant differences in the propensity of those who contribute and defect to punish. Finally, those who contribute and defect both carry out punishment against defectors. Some defectors do punish cooperators, but less often than they punish other defectors. The determinants of punishment are largely consistent across cities.

Keywords Public goods · Cooperation · Punishment

JEL Classification C91 · C92 · D90 · H41

1 Introduction

People's ability to cooperate and overcome social dilemmas, such as in public goods games, is one of the most persistent findings in the field of behavioral economics (Dawes, 1980). People consistently pay costs to contribute to their group's best interests, but cooperation is fragile to defection (Dawes & Thaler, 1988). Punishment is a key mechanism through which people sustain cooperation (Fehr & Gächter, 2000, 2002). Typically, in public goods games with punishment, participants first

✉ Talbot M. Andrews
talbot.andrews@uconn.edu

[1] Department of Political Science, University of Connecticut, 365 Fairfield Way, Unit 1024, Storrs, CT 06269-1024, USA

decide how much to contribute to the public good. After seeing the decisions of other group members, participants can pay an additional cost to impose fines on their group members (Fehr & Gächter, 2000). Punishment, or second order cooperation, facilitates the successful provision of public goods when inflicted on defectors and seems to be an evolved mechanism critical to the maintenance of cooperation (Boyd et al., 2003; Tooby et al., 2006).

However, experimental evidence finds punishment can go awry. In an impressive effort of cross-cultural data collection, Herrmann et al. (2008) conducted public goods games with punishment in 16 different cities around the world. They find some individuals engage in *antisocial punishment*, where a punisher inflicts costs on a target who contributed as much or more than the punisher to the public good. Rates of antisocial punishment vary across cultures, and antisocial punishment undermines cooperation (Herrmann et al., 2008). A reanalysis of these data examines rates of *perverse punishment*, or punishment inflicted on targets who contributed more than the group's average contribution, and finds similar results (Fu & Putterman, 2018).

Importantly, both antisocial and perverse punishment are relative, defined by the relationship between either the target and punisher's contributions or the target and group's average contributions. Rather than looking when contributing less *relative* to group members invites punishment, I instead rely on the data collected by Herrmann et al. (2008)[1] to ask: Are those who contribute less (in absolute terms) to the public good punished more often and severely? Second, are those who contribute to the public good more likely to carry out punishment than those who defect? Third, do defectors and contributors punish different types of targets? Finally, is there variation in these tendencies across cultures? Importantly, these analyses are not meant to dispute the existing work conducted using these data (Fu & Putterman, 2018; Herrmann et al., 2008). Instead, they add nuance to our understanding of the relationship between target contributions and punisher behavior.

2 Who is punished?

People who contribute more to the public good are less likely to be punished and receive smaller punishments, pooling across samples (Fig. 1). This pattern is consistent across cities: Fig. 1 plots the marginal effect of target contributions on punishment (see Tables A1 and A2 for full analyses). In no case do increased contributions lead to more punishment, though in Riyadh and Muscat there is no significant relationship between target contributions and punishment. In general, for every 1 token *more* a player contributed, they were punished between 0.05 and 0.10 tokens *less*.

This pattern is consistent when examining the probability of being punished in each city (Figure A1). In the online appendix, I present robustness checks using different model specifications further indicating those who contribute more

[1] Data was obtained from Herrmann et al. (2017).

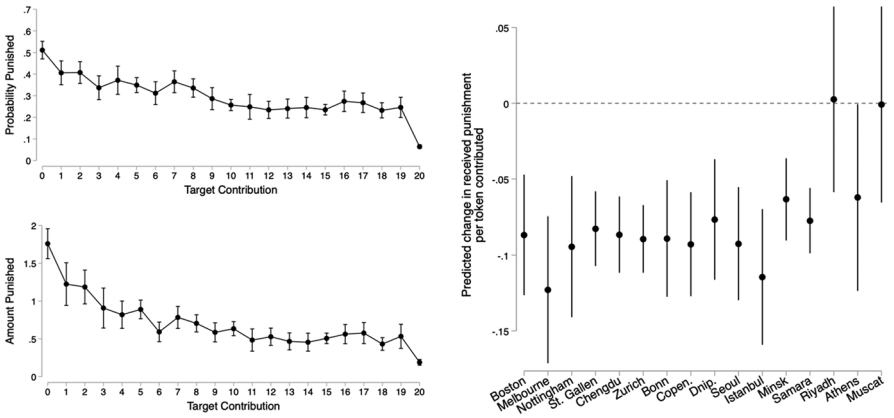

Fig. 1 The left panel shows the predicted probability someone is punished and the predicted punishment they receive by their contribution (Table A1). The right panel shows the predicted change in punishment amount for each additional token a target contributes to the public good (Table A2). For all panels, vertical bars are 95% confidence intervals clustered at the individual and group level

are generally punished less, and increased contributions do not invite increased punishment (Tables A2–A3 and Figures A2–A5).

3 The decision to punish

We might expect first-order cooperators—those who contribute to the public good—are more likely to be second order cooperators and pay the cost of punishment. Both are forms of costly cooperation (Yamagishi, 1986) that may be determined by similar underlying mechanisms (e.g., inequality aversion, Fehr & Schmidt, 1999). However, existing empirical work in the U.S. and U.K. fails to find a correlation between an individual's decision to cooperate and punish (Molleman et al., 2019; Peysakhovich et al., 2014; Weber et al., 2018).[2] Are those who contribute to the public good more willing to pay the cost of punishment across cultures?

To identify the relationship between first and second order cooperation, I regress whether someone punished and how much they punished on how much they contributed, how much the target contributed, and an interaction between the two while allowing for non-linear effects of punisher contributions (Fig. 2). There is a weak relationship between contributions and punishment, pooling across samples and across contributions of the target.

To explore differences in these results across samples, I regress whether someone punishes on their contributions in the same period, an indicator for each study location, and an interaction between the two (Fig. 2). In many cities, those who

[2] Other work has similarly failed to find correlations of behavior in different games predicted by inequality aversion (Blanco et al., 2011).

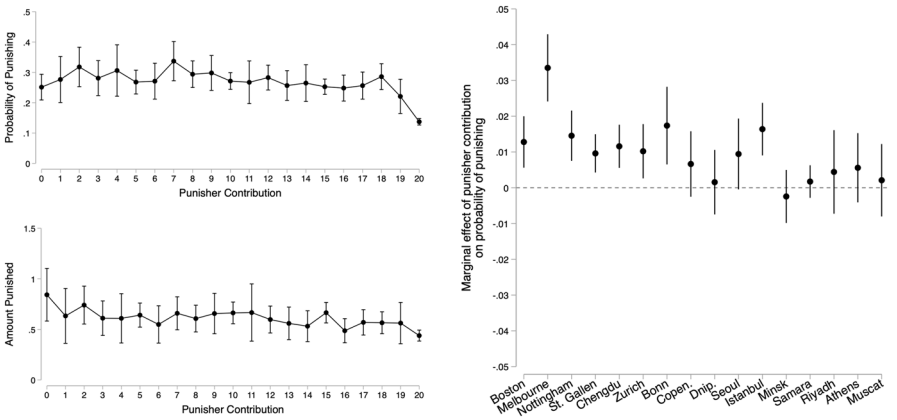

Fig. 2 The left panel shows the predicted probability someone punishes and the predicted amount of punishment over their contribution, aggregated over contributions of the target (Table A4). The right panel shows the predicted change in probability of punishing for each additional token a punisher contributes to the public good (Table A2). For all panels, vertical bars are 95% confidence intervals clustered at the individual and group level

contribute to the public good are not more likely to punish than defectors. Robustness checks with alternative model specifications are available in Figures A6–A8 and Table A2. Cooperation and punishment are not necessarily orthogonal, but these results provide additional cross-cultural evidence that first order cooperation is not a necessary condition for second order cooperation.

Do cooperators and defectors punish different types of behavior? Figure 2 aggregates over contributions of the target, which may conceal variation in whether those who contribute more punish different types of players. To identify whether this is the case, I divide contribution decisions into four categories: Defect (contribute nothing), low contribution (1–9 tokens), high contribution (10–19 tokens), and full contribution (contribute 20 tokens). I then regress whether someone punishes and how much they punish on their contribution decision, the target's decision, and an interaction between the two (Fig. 3).[3]

Defectors and full contributors alike engage in punishing defectors, though full contributors are more likely to do so. All types of players punish those who contribute nothing more frequently and severely than those who contribute everything. However, low contributors are more willing than high contributors to punish players who contribute everything. This pattern of behavior is consistent across cultures, with the exception of Riyadh and Muscat (Figures A9).

[3] Tables A5 through A8 show the percent punished and average punishment, respectively, at the intersection of each contribution level for punishers and targets, both pooled across samples and by culture. These descriptive patterns reflect the regression results.

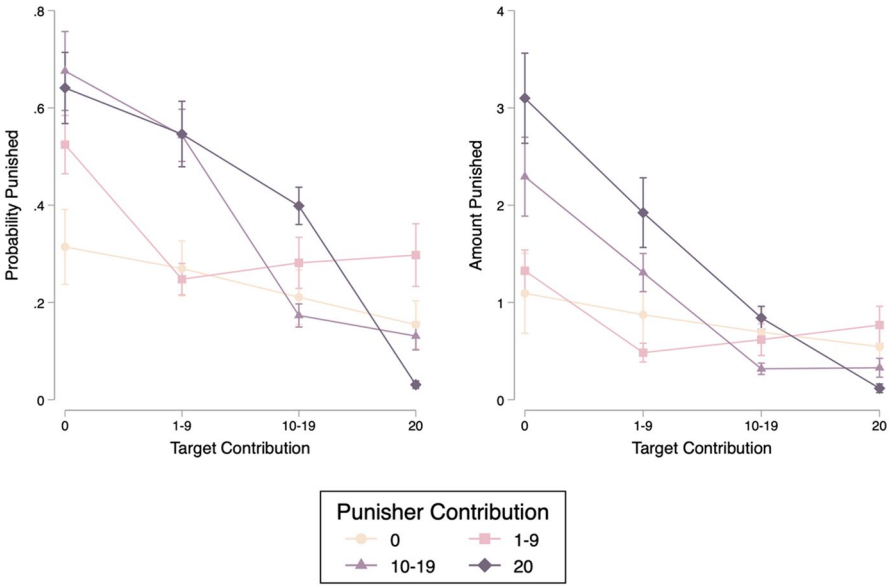

Fig. 3 The probability (left panel) and amount (right panel) players are punished at each contribution level of the punisher and target. For all panels, vertical bars are 95% confidence intervals clustered at the individual and group level. Full results are available in Table A9

4 Conclusion

Across cities, those who contribute more are punished less. While this punishment is more often carried out by those who also contribute to the public good, defectors punish one another as well. The consistency in this relationship is striking given the observed variation in contributions to the public good across cities in these data (Figure A10).

While this paper seeks to describe punishment across cultures, I leave open the question of the mechanisms driving these behaviors. First, why is there so much variation in first order cooperation across cultures (Henrich et al., 2005)? Second, why do first order defectors sometimes engage in second order cooperation? A large body of research has attempted to explain the motivations underlying conditional cooperation (Falk & Fischbacher, 2006; Fehr & Schmidt, 1999) and altruistic punishment (Fehr & Gächter, 2002), but our knowledge would benefit from future work explicitly exploring potential differences between the two.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40881-023-00157-z.

Acknowledgements I thank Scott Bokemper, Andrew W. Delton, Reuben Kline, Patrick Kraft, Yanna Krupnikov, John Barry Ryan, Christian Thöni, and Oleg Smirnov for their helpful feedback on the manuscript.

Data availability All data is available online at the following link: https://datadryad.org/resource/doi:10.5061/dryad.8730. The replication and supplementary material for the study is available at: https://doi.org/10.7910/DVN/3KC9TT.

Declarations

Conflict of interest The author declares no conflicts of interest.

References

Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior, 72*(2), 321–338. https://doi.org/10.1016/j.geb.2010.09.008

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America, 100*(6), 3531–3535. https://doi.org/10.1073/pnas.0630443100

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31*, 169–193.

Dawes, R. M., & Thaler, R. H. (1988). Anomalies: Cooperation. *Journal of Economic Perspectives, 2*(3), 187–197. https://doi.org/10.1257/jep.2.3.187

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54*(2), 293–315. https://doi.org/10.1016/J.GEB.2005.03.001

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980–994. https://doi.org/10.1257/aer.90.4.980

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137–140. https://doi.org/10.1038/415137a

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817–868.

Fu, T., & Putterman, L. (2018). When is punishment harmful to cooperation? A note on antisocial and perverse punishment. *Journal of the Economic Science Association, 4*(2), 151–164. https://doi.org/10.1007/s40881-018-0053-6

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., et al. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences, 28*(06), 795–815. https://doi.org/10.1017/S0140525X05000142

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science, 319*(5868), 1362–1367. https://doi.org/10.1126/science.1144237

Herrmann, B., Thöni, C., & Gächter, S. (2017). Data from: Antisocial punishment across societies. Dryad Digital Repository. https://doi.org/10.5061/dryad.87301

Molleman, L., Kölle, F., Starmer, C., & Gächter, S. (2019). People prefer coordinated punishment in cooperative interactions. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-019-0707-2

Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nature Communications, 5*(1), 4939. https://doi.org/10.1038/ncomms5939

Tooby, J., Cosmides, L., & Price, M. E. (2006). Cognitive adaptations for n-person exchange: The evolutionary roots of organizational behavior. *Managerial and Decision Economics, 27*(2–3), 103–129. https://doi.org/10.1002/mde.1287

Weber, T. O., Weisel, O., & Gächter, S. (2018). Dispositional free riders do not free ride on punishment. *Nature Communications, 9*(1), 1–9. https://doi.org/10.1038/s41467-018-04775-8

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51*(1), 110–116. https://doi.org/10.1037/0022-3514.51.1.110