

Punishers May Be Chosen as Providers But Not as Recipients

Yutaka Horita*

Department of Behavioral Science, Graduate School of Letters, Hokkaido University, Sapporo 060-0810, Japan.

*Author for correspondence (horita@lynx.let.hokudai.ac.jp)

The issue of evolution of punitive behavior has been a focus of recent studies of human cooperation. One of the topics for discussion in this literature is whether punishers receive benefits, on which no clear conclusion has been reached yet. We conducted a scenario experiment in which we manipulated game types and reward types, and found that punishers were chosen more frequently than non-punishers as providers of rewards, and yet, they were chosen less frequently than non-punishers as recipients of rewards. Adaptive advantages of punishers are suggested to be in their likelihood of being chosen as providers of resources, rather than as recipients of reward.

Keywords

punishment, cooperation, reputation

Introduction

Large scale human cooperation is an evolutionary puzzle. Many people anonymously donate money to various charities; participants in laboratory experiments behave in cooperative manners even in one-shot encounters. One possible answer to this puzzle may come from punishment of non-cooperators. Experimental studies show that participants use opportunities to punish non-cooperators despite the fact that such behavior is costly to themselves (e.g., Fehr & Fischbacher, 2004; Fehr & Gächter, 2002; Yamagishi, 1986). However, punishment of non-cooperators beckons another puzzle—how is costly punishment sustained (Yamagishi, 1986)? We need to specify how punishers attain sufficient adaptive advantages despite the cost of punitive behavior. For example, if punishers receive benefits which surpass the punitive cost, punishment of non-cooperators can be sustained and evolve.

However, some studies demonstrated that punishers might not receive rewards for their punitive behavior. Kiyonari and Barclay (2008) demonstrated that cooperators who expended the cost for punishing free-riders did not receive more rewards from other members in a public goods

game. Other studies also showed that those who did not use costly punishment attained a higher level of total payoff than those who provided costly punishment (Dreber, Rand, Fudenberg, & Nowak, 2008; Fehr & Rockenbach, 2003).

These findings seem to suggest that punitive behavior is fitness-reducing or maladaptive. However, there are studies demonstrating that punishment of non-cooperators may serve as a signal of trustworthiness or fairness, and punishers would be provided with more opportunities to engage in productive joint activities with other individuals (Barclay, 2006; Nelissen, 2008). In short, previous studies seemed to generate inconsistent findings concerning the personal benefits of punitive behavior, but this seeming inconsistency may simply reflects the fact that different studies focused on different types of benefits—being chosen as an interaction partner or being provided with rewards—that are generated in different types of games. We focus on the possibility that the game or the nature of interdependence determines the type of benefits generated for the punishers. Specifically, we examine participants' preferences for the punisher vis-à-vis non-punishers as game partners in various game situations. In addition, we examined how much participants would want to provide resources to the punisher or the non-punisher in those games. Specifically, five games—prisoners' dilemma game (PDG), ultimatum game (UG; Güth, Schmittberger, & Schwarze, 1982), public goods game (PGG; Yamagishi, 1986), dictator game (DG; Kahneman, Knetsch, & Thaler, 1986) and trust game (TG; Kreps, 1990)—were used for this purpose.

Methods

Fifty seven undergraduate students (40 males, 17 females) in Hokkaido University, Japan, participated. The study was conducted as a scenario experiment. Eight game situations were presented to the participants. In each scenario, depicting a particular game situation, participants were asked to imagine that they were playing an experimental game, and indicate whether they would choose a "punisher" or a "non-punisher" (to be explained below) as their game partner, and, how they would behave when their game partner was a punisher and a non-punisher in the game situation.

Before presenting the eight game situations, participants first read an "initial scenario" of either a public goods game situation in which the punisher/non-punisher suffered from non-cooperative behavior of a free-rider, or a third-party punishment game situation in which the

punisher/non-punisher was not affected by the free-rider. Which initial scenario was presented was manipulated as a between-participant factor. Across all scenarios, one of the players was depicted as a punisher who cooperated and punished a non-cooperator in the respective game.

Public Goods Game Condition

In the initial scenario of a public goods game, four players—A, B, C and D—contribute some of the endowment of 10,000 JPY to the “common fund.” Their contribution was doubled by the experimenter and was equally allocated to the four players. Players A and B (cooperators) contributed all 10,000 JPY, whereas C and D (non-cooperators) contributed nothing. The four players were further provided with an opportunity to monitor another player’s behavior, and to spend 2,000 JPY to punish the monitored player. The punished player was confiscated 10,000 JPY by the experimenter. One of the cooperators, Player A (punisher), found out that Player C did not contribute any, and spent 2,000 JPY to punish C. The other cooperator, Player B (non-punisher), also found out Player D did not contribute at all, but did not punish D.

Third-Party Punishment Condition

The four players played a dictator game as a dictator, in which they received 20,000 JPY from the experimenter, and allocated some portion of the money to a recipient. Two of the four players, A and B (cooperators), gave half of the money, 10,000 JPY, to their recipient, whereas the other two, C and D (non-cooperators), took all of 20,000 JPY and gave nothing to their recipient. The four players were further provided with an opportunity to monitor and punish another player. One of the cooperators, A (punisher), found out that C took all the money, and spent 2,000 JPY to punish C. Player B (non-punisher) also found out that D took all the money, but did not punish D.

After reading one of the two initial scenarios, all participants were provided with eight game situations. The presentation order of the eight game situations was randomized for each participant. Participants read these situations and were asked to imagine themselves in the player’s role. In each game, participants were asked which of the two players, Player A or B, they wanted to play the game with. In addition, they were asked to decide how they would behave as a game player when their game partner was a punisher (Player A) or a non-punisher (Player B).

PDG.: Two players, each received an endowment of 5,000 JPY, were asked to decide whether or not to give the money to the other. When one gave 5,000 JPY, the other received 10,000 JPY.

PGG.: Participant would be a player in a four-person game in which each of the four players received an endowment of 10,000 JPY, and decided how much of it to contribute to a “common fund.”

The total sum of the money contributed by the four members was doubled by the experimenter and equally allocated to the four players. Participants were asked to choose which group they wanted to participate in—a group in which Player A was included or Player B was included.

TG (truster’s / trustee’s role): The truster decided whether to let the experimenter assign 5,000 JPY to each of the truster and the trustee directly, or to let the trustee freely divide 20,000 JPY between the two. In the TG-truster condition, participants were asked to play the role of the truster. In the TG-trustee condition, participants played the role of the trustee.

UG (proposer’s / responder’s role): The proposer decided how to divide 20,000 JPY between him/herself and a responder. The responder decided whether to accept or reject the division. If he/she accepts the offer, both receive the amount of money according to the proposer’s decision. If he/she rejects, both receive nothing. In the UG-proposer condition, participants played the role of the proposer. In the UG-responder condition, participants played the role of the responder.

DG (dictator’s / recipient’s role): The dictator decided how to divide 20,000 JPY between him/herself and a recipient. The recipient receives the amount of money allocated by the dictator. In the DG-dictator condition, participants played the role of the dictator. In the DG-recipient condition, participants played the role of the recipient.

After participants finished answering all questions, they were individually paid and discharged.

Results

Table 1 showed the percentages of participants who wanted to play with the punisher and the mean expected earnings of the punisher and non-punisher—how much the punisher and the non-punisher would expect to earn when he/she was chosen by participants and played with them. Since the type of the initial scenario did not have a significant effect on the choice of the punisher or a significant interaction effect with punisher’s/non-punisher’s earnings in any of the eight game situations, these two conditions were pooled in the following analyses.

Participants chose the punisher more frequently than the non-punisher as their partner when he/she played the role of dictator in DG (77%), PDG player (69%), and trustee in TG (66%). The punisher was chosen more frequently than the non-punisher, though not significantly, when he/she played the role of proposer in UG (58%). In contrast, participants chose the punisher less frequently than the non-punisher as their partner when he/she played the role of responder in UG (28%), PGG player (31%), recipient in DG (33%), and truster in TG (36%). The third and the fourth columns indicate

Table 1. The Percentage of the Participants who Wanted to Play with the Punisher, the Expected Earnings of the Participants, and the Punisher's and the Non-punisher's Expected Earnings Per Game and Their Total Earnings (JPY)

Game type and the punisher/ non-punisher's role	% Punisher chosen	Participants' expected earnings		Expected earnings per game		Total earnings	
		With Punisher	With Non-punisher	Punisher	Non-punisher	Punisher	Non-punisher
Dictator in DG	76.8^a	9,351^b	7,527	10,649	12,473	7,804	2,286
PDG player	69.2	9,386	7,946	8,596	6,875	6,346	2,500
Trustee in TG	66.0	9,232	8,386	9,518	8,807	6,585	3,264
Proposer in UG	58.2	9,071	8,474	9,857	9,772	5,855	4,200
Truster in TG	36.4	8,509	9,679	7,632	7,464	2,964	4,873
Recipient in DG	32.7	11,732	12,123	8,268	7,877	2,964	4,655
PGG player	30.9	16,527	14,924	15,126	13,344	4,736	9,855
Responder in UG	27.8	10,054	11,351	9,946	8,298	2,722	5,796

a The bold number in this column indicates that the proportion was significantly different from 50% at $\alpha = .05$. b The bold numbers indicate that the punisher's earning were significantly ($\alpha = .05$) different from the non-punisher's in each pair of comparison.

that participants expected to earn more from the punisher than from the non-punisher in the games in which participants chose the punisher as a game partner. The only exception was when the punisher played the role of PGG player.

The fifth and the sixth columns indicate how much the punisher/non-punisher was expected to earn in the game. Whether the punisher was expected to earn more or less than the non-punisher depended on the game, though the difference was not significant in most games. However, the total earnings after taking into account the probability of being chosen as a game partner (and thus actually was playing the game) was largely determined by how likely the punisher/non-punisher being chosen as a game partner rather than how much they expect to earn when the game was actually played (see Columns 7 and 8). That is, the punisher's total earnings were higher than the non-punisher's only in games in which punishers were chosen more than non-punishers. The punisher's total earnings were expected to be less than the non-punisher's in the games in which he/she was avoided as a game partner, even when the punisher was expected to earn more than the non-punisher in the game.

Discussion

Punishers were chosen more often than non-punishers by participants as interaction partners in game situations where the punisher has the power to provide resources to participants—that is, in the role of dictator in DG, PDG player, trustee in TG, and proposer in UG. These findings are consistent with those reported by Barclay (2006) and Nelissen (2008). They found that punishers were entrusted more often than non-punishers by other players in TG. In contrast, punishers were chosen less frequently than non-punishers when participants had the power to determine the two players' payoffs—that is, in the role of truster in

TG, recipient in DG, PGG player and responder in UG. In those roles, the punisher was worse off than the non-punisher. As Kiyonari and Barclay (2008) claimed, the punishers are provided with fewer opportunities to become a recipient of rewards from their game partners. In short, punishing non-cooperators would be a good strategy when one is in a powerful position. In contrast, punishing non-cooperators is not a gainful strategy for those who are not likely to be in a powerful position.

Previous studies (Barclay, 2006; Kiyonari & Barclay, 2008; Nelissen, 2008) suggested the possibility that the punisher's benefits are not in the form of being rewarded by others, but being trusted by others. This study specified systematically the situations where punishers might be better off than non-punishers. Our findings further suggest the possibility that punitive behavior can evolve through individual-level selection in certain domains, especially when other individuals are looking for a game partner to play a powerful role. The reputation as a trustworthy person may be important in such a situation.

Acknowledgement

This work was supported by grants from the Japan Society for the Promotion of Science (19330139 and 08J01147). I thank Toshio Yamagishi for providing funding for the study supervising it, Victoria Yeung for her editorial help, and colleagues at Hokkaido University for helping recruitment of potential participants from their classes.

References

- Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior*, 27, 325-344. (doi: 10.1016/j.evolhumbehav.2006.01.003)
- Dreber, A., Rand, D. G., Fudenberg, D., & Nowak, M. A. (2008). Winners don't punish. *Nature*, 452,

- 348-351. (doi: 10.1038/nature06723)
- Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25, 63-87. (doi: 10.1016/S1090-5138(04)00005-4)
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137-140. (doi: 10.1038/415137a)
- Fehr, E., & Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature*, 422, 137-140. (doi: 10.1038/nature01474)
- Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental-analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 3, 367-388. (doi: 10.1016/0167-2681(82)90011-7)
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business*, 59, S285-300.
- Kiyonari, T., & Barclay, P. (2008). Cooperation in social dilemmas: free-riding may be thwarted by second-order reward than punishment. *Journal of Personality and Social Psychology*, 95, 826-842. (doi: 10.1037/a0011381)
- Kreps, D. M. (1990). Corporate culture and economic theory. In Alt, J. E., and Shepsle, K. A. (Eds), *Perspectives on positive political economy* (pp. 90-143). Cambridge: Cambridge University Press.
- Nelissen, R. M. A. (2008). The price you pay: cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior*, 29, 242-248. (doi: 10.1016/j.evolhumbehav.2008.01.001)
- Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51, 110-116. (doi: 10.1037/0022-3514.51.1.110)