

Reputational Benefit of Punishment: Comparison among the Punisher, Rewarder, and Non-Sanctioner

Hiroki Ozono*, Motoki Watabe

Graduate School of Economics, Waseda University, 1-6-1 Nishi-Waseda, Shinjuku-ku 169-8050, Japan

*Author for correspondence (hiroki.ozono@gmail.com)

Many researchers have suggested that a sanctioning system is necessary to achieve cooperation in a large society. Sanctioning others, however, is costly, raising the question of what exactly is the adaptive advantage of sanctioning. One possible answer is that sanctioners get reputational benefit. While the reputational benefits accruing to punishers and non-punishers have been compared in previous studies, in the present study we directly compared the reputational benefit of punisher, rewarder, and non-sanctioner. We conducted a scenario experiment in which participants were asked to play several games, such as the Ultimatum Game, Dictator Game, and Chicken Game with punisher, rewarder, and non-sanctioner. While in previous studies, punishers have gotten better reputational benefit as providers of resources than have non-sanctioners, we found that punishers received worse reputations than did rewarders or non-sanctioners in all games used in our experiment. These results suggest that reputational benefits change according to what kind of sanction individuals can exercise.

Keywords

punishment, cooperation, reputation, sanction, reward

Introduction

Humans can cooperate in a large society, though cooperation is costly. The existence of sanctioning systems, in which humans collectively punish non-cooperators and/or reward cooperators, is one possible reason why we can achieve large-scale cooperation, because sanctioning gives incentives for cooperation (e.g., Fehr & Fischbacher, 2004; Fehr & Gächter, 2002; Yamagishi, 1986). Sanctioning raises another problem, however, because it is costly in itself: What is the adaptive advantage of sanctioning behavior?

One possible answer is the reputational benefit to sanctioners; that is, sanctioners receive improved

reputation and thereby benefit in other social situations. The results of experiments to assess this possibility, however, have been inconsistent. Some studies have shown that punishers were thought of as trustworthy and fair, meaning that they might get reputational benefit by being chosen as interaction partners (Barclay, 2006; Nelissen, 2008). However, Kiyonari and Barclay (2008) showed that punishers did not receive more reward than non-punishers. Horita (2010) tried to explain these results consistently. In his scenario experiment, participants were asked to imagine that they played experimental games, such as the Ultimatum Game (UG; Güth, Schmittberger, & Schwarze, 1982) and the Dictator Game (DG; Kahneman, Knetsch, & Thaler, 1986) with a punisher or a non-punisher, and were asked to choose one or the other as their game partner. Horita's results showed that punishers were chosen more frequently than non-punishers as providers of resources but less frequently as recipients of resources. He concluded that punishers get an adaptive benefit by being chosen as providers rather than as recipients.

On the basis of Horita (2010), we investigated two unstudied issues in our experiment. First, we suggested another possible reputational benefit for the punisher, that is, the reputation for "toughness." Punishing others is the behavior of reducing their resources. Punishment can be motivated by anger, so punishers may develop more of a reputation for being strong, tough, and/or short-tempered. In turn, this may reduce the risk that punishers will be exploited by others. The adaptive function of toughness was discussed in Culture of honor (Nisbett & Cohen, 1996), which explained why white Americans in the South tend to rely on their sense of personal honor and reputation for toughness more than do those in the North. Nisbett and Cohen concluded that livestock farming culture in the South is the origin of this tendency, because the reputation for toughness is important to prevent others from stealing one's livestock. In human evolution, the reputation for toughness may have played an important role. Punishing others probably had the function of raising people's reputation for toughness. In our study, participants were asked to play the Chicken Game (CG; Rapoport & Chammah, 1966), in which those who have the reputation for toughness can take advantage of others; in this way, we tried to show that punishers get an adaptive benefit from their reputation for toughness.

Second, we compared the reputational benefits of punishers and rewarders directly. Most previous studies, including Horita (2010), compare the reputational benefit of punishers and non-punishers, but sanction does not encompass only punishment

for non-cooperators; it may include rewards for cooperators. Recent studies reveal that rewarding cooperators contributes to the emergence of cooperation in a situation (Rand, Dreber, Ellingsen, Fudenberg, & Nowak, 2009). Comparisons of punisher, rewarder, and non-sanctioner (one who neither rewards nor punishes), should be fruitful for the understanding of human cooperation.

Method

Participants

Fifty Japanese undergraduate students (26 women and 24 men, mean age = 20.2 years) participated.

Procedure

The study was conducted as a scenario experiment. Participants were asked to imagine that they were playing five games in the roles of player (CG), proposer and responder (UG), and dictator and recipient (DG), with five imagined partners. Participants were told that the five partners had played the Public Goods Game (PGG) and Sanction Game (SG) among them beforehand. The structure of PGG and SG, and the behaviors of the five partners described to participants were as described below.

In this PGG, seven players, A, B, C, D, E, F, and G, participated. Players B, C, D, E, and F were the five imagined partner of participants. In the PGG, seven players were first given 10,000 JPY, and decided how much money they would each contribute to the “common fund.” Their contribution was doubled by the experimenter and shared equally among all players. In the scenario, Player A contributed 10,000 JPY, players B, C, D, E, and F contributed 5,000 JPY, and player G contributed nothing. After the PGG, the Sanction Game (SG) was played, in which each player had a chance to reward or punish another player. First, the contribution of each player was revealed, and each player could then choose who would be rewarded or punished. If one player spent 1,000 JPY to reward another player, that player would be awarded 10,000 JPY by the experimenter. If one player spent 1,000 JPY to punish another player, that player was penalized 10,000 JPY by the experimenter. Players could spend any amount to reward or punish. Thus, all players decided who should be rewarded or punished and how much money they would spend on sanction, but each player could choose only one other player to reward or punish. In addition, players were told before the decision that one decision was randomly set by the experimenter. This operation was performed so that the expectation of each decision might not influence another.

Participants were told the decisions in the SG phase of the five imagined partners, player B, C, D, E, and F, who all contributed 5,000 JPY in the PGG. Players B and D rewarded player A, who was the high contributor. Player B (moderate rewarder)

spent 500 JPY and Player D (high rewarder) spent 1,000 JPY. Players C and E punished player G, who was the zero contributor. Player C (moderate punisher) spent 500 JPY and player F (high punisher) spent 1,000 JPY. Player F (non-sanctioner) did not reward or punish others.¹

After reading about the behavior of the five partners, participants chose with which partner they wanted to play in each game on a seven-point Likert-type scale (a rating of 1 indicated “never want to play” and a rating of 7 indicated “want to play very much”). They also decided their own behaviors in each game and rated their expectations of their partners’ behavior in each game. The rules of the five games and the roles of participants are below.

CG (Player)

Each player chooses whether s/he pushes her/his button. When both choose “do not push,” both get 10,000 JPY each. When one chooses “push” and the other chooses “do not push,” the “push” player gets 15,000 JPY and the “do not push” player gets 5,000 JPY. When both choose “push,” both get nothing. “Push” can be regarded as the “tough choice” and “do not push” as a “chicken choice” in this situation. Each participant decided whether s/he would push or not push and inferred the probability that her/his partner would choose “push.”

UG (Proposer/Responder)

The proposer makes an offer regarding how to divide 20,000 JPY between the responder and her/himself. The responder has a right to accept or reject the offer. When s/he accepts, both of them get money according to the offer of the proposer. When s/he rejects the offer, both of them get nothing. In the UG proposer condition, the participants decided the offer and inferred the minimum offer acceptable to the partner (responders). In the UG responder condition, the participants decided the minimum acceptable offer and the expectation of the amount of the offer by the partner (proposer).

DG (Dictator/Recipient)

The dictator decides how to divide 20,000 JPY between the recipient and her/himself. Unlike in UG, the recipient does not have a right to reject the offer; both participants receive money according to the offer by the dictator. In the DG dictator condition, the participant decided the offer. In the DG recipient condition, the participant inferred the expectation for the level of the offer by each partner (dictator).

¹ The degree of sanction (high or moderate) was controlled in the experiment, but we do not refer to the results of this factor in this article because the description of the results would become complicated if we did. We should note that the conclusion does not change whether or not we analyze this factor. In the analysis, the results for Players B and D (rewarders) and Players C and E (punishers) were integrated.

Table 1. Means of Measured Variables and Results of Multiple Comparisons in Each Variable

Roles of partners	CG player			Recipient in DG		Dictator in DG	
	Preference	Participant's choice (0: do not push, 1: push)	Expectation that partner will choose "push" (%)	Preference	Division for partner	Preference	Expectation of division for partner
Punisher	3.13 ^b (0.17)	.29(0.05)	47.5 ^a (3.28)	3.26 ^b (0.16)	7830(471)	3.03 ^b (0.16)	11920 ^a (452)
Rewarder	4.02 ^a (0.12)	.35(0.05)	37.3(2.78)	3.96 ^a (0.14)	8722(848)	4.07 ^a (0.12)	10073 ^b (299)
Non-sanctioner	4.00 ^a (0.24)	.44(0.07)	35.6 ^b (4.15)	4.06 ^a (0.25)	7080(495)	4.02 ^a (0.25)	10880(577)
<i>F</i> (2, 98)	5.44	2.16	3.48	4.02	2.45	7.65	3.68

Roles of partners	Responder in UG			Proposer in UG		
	Preference	Offered division for partner	Expectation of minimal acceptable division	Preference	Expectation of offered division for partner	Minimum acceptable division for participant
Punisher	3.10 ^b (0.18)	10200 ^a (222)	9244(231)	2.98 ^b (0.17)	11579 ^a (319)	6638(471)
Rewarder	4.21 ^a (0.13)	9323 ^b (268)	7526(358)	4.28 ^a (0.12)	9942 ^b (191)	7271(444)
Non-sanctioner	3.98 ^a (0.25)	9182 ^b (329)	9302(1690)	4.10 ^a (0.25)	10330 ^b (493)	6750(472)
<i>F</i> (2, 98)	6.63	4.20	0.99	10.27	5.14	2.60

Note. Standard errors are in parentheses. "a" is significantly larger than "b" ($p < .05$; multiple comparison using Ryan's method).

Results and Discussion

The means of the measured variables are shown in Table 1. We performed one-way factorial ANOVAs with each variable; multiple comparisons were conducted.

There were two major findings in our results. First, we found that punishers got a reputation for "toughness." In CG, punishers were thought more likely to push the button (the "tough" choice) than were non-sanctioners (see "expectation that partner will choose 'push'" in CG Player). This suggests that punishers may get an adaptive benefit in situations where toughness plays an important role, such as a resource-shortage situation in which the risk of being exploited is large. Punishment behavior possibly evolved as an adaptation to such an environment, though there should be further research into whether human beings need such a reputation in an environment of evolutionary adaptedness.

Our second finding was that punishers seemed to acquire a worse reputation than did rewarders or non-sanctioners, across all games. While Horita (2010) showed that a punisher was more desired than a non-punisher as a provider of resources, in other words in roles such as proposer in UG or dictator in DG, we showed that the punisher was less liked even in these games (see "preference" in UG Proposer and in DG Dictator). In addition, we showed that punishers were not expected to divide resources as fairly as rewarders and non-sanctioners in UG and DG (see "expectation of division for partner" in UG Proposer and DG Dictator), though punishers were thought of as fair in the study by Barclay (2006). Why have these inconsistent results been observed? In the experiments by Horita (2010) and Barclay (2006), partners had only two options: punish or do not punish. In contrast, partners in our experiment had three options: punish, reward, or neither punish nor reward. These differences

might influence the impression of partners. When there are three options, partners have a chance to reward high cooperators. Nevertheless, partners who chose to punish non-cooperators were thought to be spiteful and unjust. In other words, punishing others might be justified when and only when they do not have a chance to reward cooperators. These results suggested that reputational benefit changes according to what kind of sanction can be exercised. Furthermore, other factors, such as whether the sanctioning system obtains a social consensus or whether the procedure for the sanction is justified, may influence the reputation of sanctioners. In future research, it will be necessary to examine differences in reputational benefits according to what kind of sanctioning systems individuals can exercise. In addition, this study was conducted as a scenario experiment, so participants did not have any incentives in their decision-making. In the future, it is necessary to conduct an experiment in which actual money is used in order to examine the actual behaviors of participants when they have incentives.

References

- Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior*, 27, 325–344. (doi: 10.1016/j.evolhumbehav.2006.01.003)
- 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)
- 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)

- Horita, Y. (2010). Punishers may be chosen as providers but not as recipients. *Letters on Evolutionary Behavioral Science*, 1, 6–9. (doi: 10.5178/lebs.2010.2)
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business*, 59, S285–300. (doi: 10.1086/296367)
- Kiyonari, T., & Barclay, P. (2008). Cooperation in social dilemmas: Free-riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology*, 95, 826–842. (doi: 10.1037/a0011381)
- 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)
- Nisbett, E. R., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the South*. Boulder, CO: Westview Press.
- Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science*, 325, 1272–1275. (doi: 10.1126/science.1177418)
- Rapoport, A., & Chammah, A. M. (1966). The game of chicken. *American Behavioral Scientist*, 10, 10–28. (doi: 10.1177/000276426601000303)
- 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)