LEBS
Vol. 11 No. 1 (2020) 6–9
Letters on Evolutionary Behavioral Science

The Evaluation of Second- and Third-Party Punishers

Nobuhiro Mifune[1,*], Yang Li[2], Narumi Okuda[3]

[1]Kochi University of Technology

[2]The University of Melbourne

[3]Fujikin CO., LTD.

[*]Author for correspondence (n.mifune@gmail.com)

Although punishment can promote cooperative behavior, the evolution of punishment requires benefits which override the cost. One possible source of the benefit of punishing uncooperative behavior is obtaining a positive evaluation. This study compares evaluations of punishers and non-punishers. Two hundred and thirty-four undergraduate students participated in two studies. Study 1 revealed that, in the public goods game, punishers were not positively evaluated, while punishers were positively evaluated in the third-party punishment game. In Study 2, where the non-cooperator was a participant of a public goods game, we manipulated the punishers participation in the game. The results showed that punishers received no positive evaluations, regardless of their participation in the game, indicating that negative evaluation may not be a reaction toward aggression with retaliatory intentions.

Keywords
public goods game, third-party punishment, punishment

Introduction

Humans cooperate with non-familiar others, which is a unique characteristic. The evolutionary mechanism of this unique cooperation among *Homo sapiens* has resulted in debate among fields such as evolutionary biology, anthropology, and economics (e.g., Bernhard et al., 2006; Nowak & Sigmund, 1998; Tomasello et al., 2012). One explanation of this cooperation with non-familiar others is punishment directed toward non-cooperators (Fehr & Gächter, 2002; Yamagishi, 1986). Rather than the benefits of non-cooperation being reduced by punishment, it is preferable to cooperate; thus, cooperation can evolve (Boyd et al., 2003). Since punishment raises the overall cooperation level within the group, at the detriment to the punisher, it creates a second-order dilemma where individuals are induced to free-ride on others' second-order cooperation (i.e., punishment). The resolution of the debate concerning the evolution of cooperation and punishment requires resolution of the second-order dilemma.

One important aspect in understanding the evolution of punishment is evaluations of punishers by others (Raihani & Bshary, 2015a). If the punishment results in a positive image of the punisher and increases the possibility of cooperation with others, punishments can evolve. The evaluation of punishers has been primarily studied using two paradigms: public goods with punishment (PGP) and third-party punishment (TPP) games.

The PGP game groups a participant with others and the participant decides how much to contribute to the group from his/her own endowment. It is profitable for individuals not to contribute money to the group; however, the income of the entire group increases with contributions, resulting in distributed benefit. Following the contribution stage, participants decide whether to deduct money from other players' payoffs (punishment). Previous studies have found that punishers in the PGP game are not evaluated positively (Kiyonari & Barclay, 2008), are not rewarded by others (Kiyonari & Barclay, 2008), and are not given preference as potential partners in economic games (Ozono & Watabe, 2012; but see also Barclay, 2006).

In the TPP game, three participants play the roles of allocator, recipient, and observer. In the TPP, the allocator and recipient first play a dictator game: Allocators receive an endowment from the experimenter to allocate between themselves and the recipient. The observer is informed of the allocation and decides to what extent (if any) to deduct the payoff from the allocator (punishment). In previous studies using the TPP game, observers who implement punishment are more positively evaluated than those who do not punish (Nelissen, 2008), are more likely to be chosen as a partner of a game (Nelissen, 2008), and are more likely to be rewarded (Raihani & Bshary, 2015b).

It has been noted that the reason punishers in the PGP game are negatively evaluated is that they are perceived as retaliatory or aggressive figures, while this is not the case in the TPP game where they are positively evaluated (Raihani & Bshary, 2015a, 2019). Cooperators in the PGP receive less payment when there is a non-cooperator in the group compared to when everyone cooperates. It is common for punishers in the PGP game to contribute to the public good, resulting in the punishment as a form of revenge for the loss of benefits due to defection; thus, the punisher is less likely to receive a positive evaluation. Since punishers in the TPP game are not perceived as victims of a "selfish" allocation, the punishment is less likely to be perceived as vengeful, and may result in more positive evaluations.

Study 1

Study 1 tested the hypothesis that, in comparison to non-punishers, punishers in the PGP game are more negatively evaluated, while punishers in the TPP game are more positively evaluated. Although Horita (2010) reported that

doi: 10.5178/lebs.2020.72
Received 23 January 2020.
Accepted 02 February 2020.
Published online 06 February 2020.

the likelihood of punishers in the PGP game being chosen as a game partner did not differ from those in the TPP game, to our knowledge, no study has directly compared the evaluations of punishers in the two games. This study compared the differences in evaluations of punishers between the two games using the same participants.

Method

One-hundred and seven undergraduate students (44 male, 62 females, 1 unknown) participated in the study. Each participant visited the laboratory individually and completed a survey, receiving a payment of 1000 JPY (approximately 10 USD). The survey consisted of questions from multiple study projects with different purposes; however, only the questions related to this study are reported here. The relevant questions were presented at the beginning of the survey.

The study utilized a 2 (game type: PGP vs. TPP) by 2 (target: punisher vs. non-punisher) within-subjects experimental design. The dependent variable was the evaluation of the target figure. The sequence of the game was counter-balanced, while in each game punishers was always evaluated before non-punishers.

In each game scenario, participants were asked to imagine the described game and then evaluate the target figure the scenarios described. In the PGP, four people participated in a two-stage game comprising an investment stage and a deduction stage. Each person was provided an endowment of 1000 JPY, from which they could invest any amount to the public goods. The experimenter aggregated and doubled the investment in the public goods, and then equally distributed the final amount among the four participants. At the end of the investment stage, each of the four participants received a share from the public goods and whatever amount they had kept for themselves rather than investing in the public goods. In this scenario, participant A kept the full 1000 JPY endowment for themselves, while participant B, C, and D all invested the entire 1000 JPY endowment in the public goods. Consequently, at the end of the investment stage, participant A received a total of 2500 JPY, while participants B, C, and D each received 1500 JPY. The game then proceeded to the deduction stage, where another endowment of 500 JPY was provided to each participant. The four participants were informed that they could spend any amount from the deduction endowment to reduce another person's earnings for 3 times of the cost. Participant B decided to use the entire 500 JPY to deduct 1500 JPY from participant A, while participants A, C, and D did not use any of the endowment for deduction. At the end of the public goods game, participant A received 1500 JPY, participant B received 1500 JPY, participant C received 2000 JPY, and participant D received 2000 JPY. The participants studied the resultant public goods scenario and were asked for their evaluation of participant B, who elected to significantly reduce participant A's earnings and participant C who did not pay in the deduction stage.

In the TPP game, participants A and B were paired. Participant A received an endowment of 1000 JPY to be allocated freely between the two participants. Participant A decided to keep the full endowment; thus, participant B received nothing. Participant C observed the allocation and received a deduction endowment of 500 JPY, any amount from which could reduce participant A's earnings for twice of the cost. Participant C decided to spend all the deduction endowment to reduce 1000 JPY from participant A's earnings. Contrastingly, in another third-party punishment scenario, where participants D and E were paired and D also kept the full 1000 JPY endowment, a third-party observer, participant F, decided to keep the 500 JPY endowment rather than deduct anything from D.

After each scenario of the PGP and TPP games, the participants rated those who punished (participant B in PGP and participant C in TPP) as well as those who did not punish (participant C in PGP and participant F in TPP). Image evaluation items from Kiyonari and Barclay (2008) were used for the evaluation, where six items involving trustworthiness, cooperativeness, generosity, likability, goodness, and dependability are rated on a 9-point Likert scale (1 - strongly disagree to 9 - strongly agree).

Results

The six evaluation items were highly consistent under each condition (αs > .86); therefore, the following analysis used the average score of the six items as a dependent variable. Figure 1 illustrates the mean scores of the evaluations in each condition. An ANOVA using target and game type as independent variables found a main effect of game type ($F(1, 106) = 7.36$, $p < .01$) and an interaction between the two, ($F(1, 106) = 54.57$, $p < .01$), but no main effect of

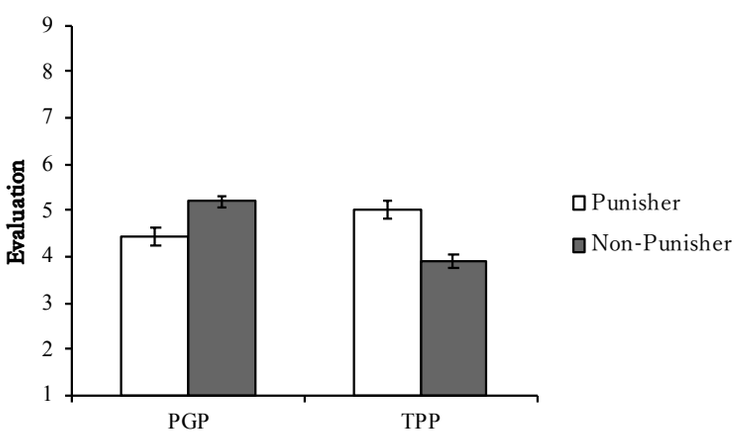

Figure 1. Average evaluation scores in PGP and TPP in Study 1.

Note. Error bars reflect standard errors *PGP: Public Goods with Punishment, TPP: Third Party Punishment

target ($F(1, 106) = 1.25$, $p = .27$). The simple main effect analysis showed that punishers were more negatively evaluated than non-punishers in the public goods game ($p < .01$, adjusted by Holm method). In the third-party punishment game, punishers received more positive evaluations ($p < .01$, adjusted by Holm method).

Study 2

The results from Study 1 are consistent with previous findings that punishers in the PGP game are evaluated negatively while punishers in the TPP game are evaluated more positively (e.g., Kiyonari & Barclay, 2008; Nelissen, 2008). The results indicate that the negative evaluations are due to punishers in the PGP game potentially being considered to be retaliators since the punisher was the victim of a defection. Despite these findings, in this case, there was a confound of game type and participation of public goods. If the negative evaluation toward punishers in the PGP is because the punishment was perceived as retaliation, it is possible that punishers who have not participated in the public goods game are not necessarily evaluated negatively. In Study 2, we focused on the public goods game and manipulated whether the potential punisher had participated in the public goods game.

Method

One hundred and twenty-seven undergraduates participated in Study 2 (76 males, 51 females). Participants were recruited from a psychology class. After reading the consent form on the front page, only those who agreed to participate proceeded to the survey.

Study 2 considered whether the punisher played the public goods game and used a 2 (player vs. non-player) by 2 (punisher vs. non-punisher) within-subjects design. Consistent with Study 1, the dependent variable was the evaluation of the targets. Both the game type and target were counter balanced in their presentation order.

In the player condition, participants read a scenario in which four people played a public goods game, which was identical to Study 1. One of the four people was a non-cooperator, while the other three were cooperators. In the second stage, one of the cooperators in the public goods game punished the non-cooperator, while the others did not. After reading the scenario, the participants evaluated the punisher and the non-punisher.

In the non-player condition, the study presented a scenario of six people to the participants. Four of the six people were randomly selected to participate in the first stage (public goods game), while the remaining two people participated in the second stage (punishment stage). The public goods game scenario was consistent with both Study 1 and the player condition, where one person did not cooperate while the other three people cooperated. In the second stage, the two people who did not participate in the public goods game received a deduction endowment of 500 JPY. One of the two people used up the full endowment to deduct 1500 JPY from the non-cooperator, while the other kept the endowment for their own use. After reading the scenario, the participants evaluated the punisher and the non-punisher.

At the end of the survey, the participants answered several demographic questions. The study was completed in about 15 minutes.

Results

As for Study 1, the six evaluation items were highly consistent in Study 2 (αs >.90); thus, we used the average score as the dependent variable. Figure 2 shows the average scores for each condition. An ANOVA confirmed significant main effects for both player ($F(1, 126) = 5.68$, $p < .05$) and punisher ($F(1, 126) = 34.68$, $p < .01$), while no interaction was found ($F(1, 126) = 0.04$, $p = .85$). Similar to Study 1, non-punishers were perceived more positively in the player condition ($p < .01$, adjusted by Holm method), and this was also observed in non-player condition ($p < .01$, adjusted by Holm method).

Discussion

Both Study 1 and Study 2 revealed that participants perceived punishers who participated in the public goods game more negatively in comparison to non-punishers. Meanwhile, as has been found in the previous literature (Nelissen, 2008), Study 1 also demonstrated that punishers in a third-party punishment game were rated more positively than non-punishers. These results indicate that punishers in the public goods game were perceived as seeking revenge because they suffered from defection, and such retaliation were evaluated negatively

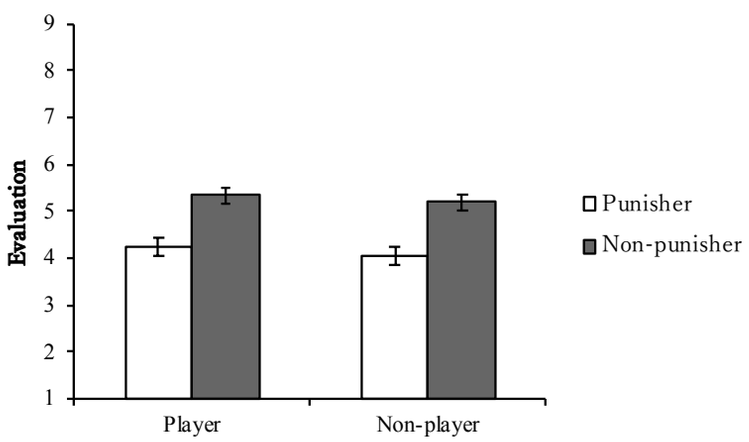

Figure 2. Average evaluation scores of player and non-player in PGP in Study 2.

Note. Error bars reflect standard errors

by others. Study 2 found that a punisher who had been a third-party observer in the public goods game received an equally negative evaluation as those punishers who experienced defection in the public goods game, indicating that the negative evaluations of the public goods game punishers might not be caused by the retaliatory aspect of the punishment. These results imply that the different evaluations of punishers in the PGP and TPP reported in previous studies may not only be a reaction to the perceived retaliation of the punisher. Rather, we propose that the legitimacy of the punishment in PGP may be perceived differently among evaluators, reflecting their inference of the intention of non-cooperation. Further investigation for the topic is required.

While punishment may need to be perceived as legitimate for the punisher to receive positive evaluation (Raihani & Bshary, 2019), further research is required to investigate why the punishment of non-cooperators by a third-party observer is not considered as legitimate in the public goods game. Since punitive strategies evolve only when the cost of punishment is compensated by some form of benefit, it should be investigated whether punishers gain rewards other than evaluations and whether they are more likely to be chosen as game partners.

References

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

Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group affiliation and altruistic norm enforcement. American Economic Review, 96, 217-221. https://doi.org/10.1257/000282806777212594

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, 3531-3535. https://doi.org/10.1073/pnas.0630443100

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

Horita, Y. (2010). Punishers may be chosen as providers but not as recipients. Letters on Evolutionary Behavioral Science, 1, 6-9. https://doi.org/10.5178/lebs.2010.2

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. https://doi.org/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. https://doi.org/10.1016/j.evolhumbehav.2008.01.001

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. Nature, 393, 573-577. https://doi.org/10.1038/31225

Ozono, H., & Watabe, M. (2012). Reputational benefit of punishment: comparison among the punisher, rewarder, and non-sanctioner. Letters on Evolutionary Behavioral Science, 3, 21-24. https://doi.org/10.5178/lebs.2012.22

Raihani, N. J., & Bshary, R. (2015a). The reputation of punishers. Trends in Ecology & Evolution, 30, 98-103. https://doi.org/10.1016/j.tree.2014.12.003

Raihani, N. J., & Bshary, R. (2015b). Third-party punishers are rewarded, but third-party helpers even more so. Evolution, 69, 993-1003. https://doi.org/10.1111/evo.12637

Raihani, N., & Bshary, R. (2019). Punishment: one tool, many uses. Evolutionary Human Sciences, 1, e12. https://doi.org/10.1017/ehs.2019.12

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation: the interdependence hypothesis. Current Anthropology, 53, 673-692. https://doi.org/10.1086/668207

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