Self-Punishment Promotes Forgiveness in the Direct and Indirect Reciprocity Contexts

Psychological Reports
2017, Vol. 120(3) 408–422

DOI: 10.1177/0033294117697087
journals.sagepub.com/home/prx

Ruida Zhu, Xueyi Shen, Honghong Tang, Peixia Ye, and Huagen Wang
State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China; Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, China

Xiaoqin Mai
Department of Psychology, Renmin University of China, Beijing, China

Chao Liu
State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China; Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, China

Abstract

Most previous studies regarding self-punishment have focused on the correlation between moral emotion and self-punishment. Only a few studies have attempted to understand self-punishment from the perspective of seeking forgiveness, and no study has yet directly tested whether wrongdoers' self-punishment promotes others to forgive the wrongdoers. In three studies, the participants judged the wrongdoers' self-punishment behaviors following an unfair allocation and reported the extent to which they forgave the wrongdoers. The results demonstrated that self-punishment did promote forgiveness in both the direct (Studies 1 and 2) and indirect reciprocity

Ruida Zhu and Xueyi Shen contributed equally to this work.

Corresponding Authors:
Xiaoqin Mai, Department of Psychology, Renmin University of China, No. 59 Zhongguancun Street, Beijing 100872, China.
Email: maixq@ruc.edu.cn
Chao Liu, State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, No. 19 Xinjiekouwai Street, Beijing 100875, China.
Email: liuchao@bnu.edu.cn

(Study 3) contexts. Consistent with costly signaling theory, the costlier the self-punishment was, the stronger the effect it had on forgiveness. Moreover, communicative self-punishment had a better effect than silent self-punishment when the cost was relatively high in the direct-reciprocity studies. These findings can guide us regarding how to address a damaged relationship via self-punishment when compensation is not feasible or acceptable.

Keywords
Self-punishment, forgiveness, cost, communicative, silent

Introduction

After committing a transgression, wrongdoers sometimes punish themselves by inflicting physical pain or economic loss on themselves (Bastian, Jetten, & Fasoli, 2011; Inbar, Pizarro, Gilovich, & Ariely, 2013; Nelissen, 2012; Nelissen & Zeelenberg, 2009; Ohtsubo et al., 2014; Tanaka, Yagi, Komiya, Mifune, & Ohtsubo, 2015; Watanabe & Ohtsubo, 2012). This interesting phenomenon, named self-punishment, is widespread in social, economic and religious activities (Leff, 1999; Wallington, 1973).

Most previous studies have explained wrongdoers' self-punishment from the point of view of emotion. After violating moral norms, wrongdoers suffer from guilt and shame. Their moral belief that transgression should be balanced by punishment prompts them to alleviate their negative emotions by punishing themselves (Inbar et al., 2013). It was revealed that guiltier wrongdoers tend to engage in severer self-punishment (Gintis, Smith, & Bowles, 2001; Nelissen, 2012; Nelissen & Zeelenberg, 2009; Tanaka et al., 2015; Watanabe & Ohtsubo, 2012), and severer self-punishment relieves more guilt (Bastian et al., 2011; Inbar et al., 2013).

A few studies have tried to understand self-punishment from the point of view of seeking forgiveness (Nelissen, 2012; Watanabe & Ohtsubo, 2012). After committing a transgression, wrongdoers face the risk of revenge from the victim and being excluded from reciprocity by their group (Cashdan, 1980). Therefore, wrongdoers may strategically use self-punishment to signal their remorse and restore the damaged relationship. Nelissen (2012) found that wrongdoers were more inclined to punish themselves in front of the victim—who was more likely to seek revenge—than in front of a bystander. These results imply that wrongdoers attempt to seek forgiveness from others by exerting punishment on themselves. However, others' attitude and responses toward wrongdoers' self-punishment have not yet been studied. We explore whether and how wrongdoer's self-punishment affects others' forgiveness in this article.

The cost of self-punishment is a possible factor that may influence forgiveness. Based on costly signaling theory, receivers perceive costlier signals to be more credible and reliable (Gintis et al., 2001; Roberts, 1998; Smith & Bird,

2000; Sosis, 2000; Zahavi, 1977). Previous studies found when the wrongdoer was willing to bear more cost to apologize or compensate, the victim was more likely to feel their sincerity (Eaton, Struthers, & Santelli, 2006; Gauché & Mullet, 2005; Ohtsubo & Watanabe, 2009; Ohtsubo et al., 2012). It is possible that people consider self-punishment as a sincere expression of remorse due to the cost and forgive the wrongdoer. Thus, it is likely that the costlier a self-punishment is, the more likely others will forgive the wrongdoer.

Another factor that may influence forgiveness is whether the wrongdoer actively informs others about his self-punishment. In some cases, wrongdoers punish themselves to relieve their own guilt and shame; thus, they do not deliberately inform others about their self-punishment behavior and may even undertake the behavior in private (Bastian et al., 2011; Inbar et al., 2013). This type of self-punishment is named silent self-punishment. In other cases, to seek forgiveness from others, wrongdoers punish themselves in public and attempt to inform others about their self-punishment (Nelissen, 2012). This behavior is called communicative self-punishment. Whether communicative self-punishment can result in the desired consequence is uncertain. It has been found that people care about the motives underlying others' behaviors (Berman, Levine, Barasch, & Small, 2015). When wrongdoers attempt to broadcast their self-punishment, others may think that the wrongdoers are simply afraid of revenge or exclusion from reciprocity and thus only pretending to be regretful. Therefore, the effect of communicative self-punishment on forgiveness could be worse than that of silent self-punishment and even worse than doing nothing. In contrast, it is also possible that communicative self-punishment has a more positive effect on forgiveness compared to silent self-punishment. Self-punishers who follow social norms usually are good potential cooperators in cooperation (Baumard, André, & Sperber, 2013; Trivers, 1971). However, if a wrongdoer does not inform others about their self-punishment, people may mistakenly think that the wrongdoer does not regret committing the transgression and incur an unnecessary cost to punish the wrongdoer. Therefore, when people accidently find wrongdoers punishing themselves silently, they may blame the wrongdoers for not informing others and inducing unnecessary cost. Considering the two opposite opinions, it is worth testing the effect of communicative and silent self-punishment. Together, the first aim of our studies was to investigate how the cost and type of self-punishment affect forgiveness.

Many studies about forgiveness focus on victims' attitudes towards wrongdoers (for a review, see McCullough, Kurzban, & Tabak, 2013). From the perspective of indirect reciprocity theory, it is worth noting that bystanders judge wrongdoers' behavior as well, and many of them are willing to incur some cost to punish wrongdoers (Fehr & Fischbacher, 2004). Therefore, it is necessary to also study bystanders' attitudes and behaviors toward wrongdoers. So far, very few studies have investigated the effect of self-punishment in the indirect reciprocity context (Tanaka, Ohtsuki, & Ohtsubo, 2016). Thus, the

second aim of our study was to investigate the effect of self-punishment on forgiveness in both the direct reciprocity (the victim and wrongdoer, Studies 1 and 2) and indirect reciprocity (the bystander and wrongdoer, Study 3) contexts.

Study 1

Method

Participants and design. A total of 78 undergraduate students participated in the study for a monetary payment. The participants performed the task on the computer alone in individual rooms. The study had a within-subject design. This study was approved by the institutional review board of the Beijing Normal University.

Procedure. Participants received instructions about an allocation game (adapted from Ohtsubo & Watanabe, 2009, Experiment 3). It was explained that there were an allocator and a receiver in the game. The allocator was required to decide how to distribute 10 Chinese yuan between the receiver and himself. Nine allocation plans were possible, which gave the receiver one to nine Chinese yuan in steps of one. However, only two allocation plans, which were randomly picked from the nine potential plans by the computer program, were offered to the allocator. Thus, the allocator could only choose one plan from the two offered plans. The receiver had no choice but to accept the allocation. Both allocator and receiver knew the rules of the game above. However, the receiver did not know which two allocation plans were offered to the allocator (this manipulation blurred the intention of the allocator and made the allocator's consequent behavior reasonable). The participants were asked to imagine that they played the role of the receiver and judge the behavior of the five different allocators in the following tasks. Then, participants completed a test regarding their comprehension of the instructions. Only when participants correctly answered all the questions were their data used in the analyses.

Following the comprehension test, the participant imagined that he or she and five different allocators were seated in separated rooms. The participant played the allocation game once with each of allocators. It turned out that all allocators chose a plan in which the participant received two Chinese yuan, whereas the allocator obtained eight. However, different actions were taken by different allocators after the unfair allocation as follows: allocator A did nothing (the control condition); allocator B abandoned 2 Chinese yuan and informed the participant how much money he or she had abandoned (the low-cost communicative self-punishment condition, LCCSP); allocator C abandoned 6 Chinese yuan and informed the participant how much money he or she had abandoned (the high-cost communicative self-punishment condition, HCCSP). Allocator D abandoned 2 Chinese yuan silently, and the participant learned about it

accidently (the low-cost silent self-punishment condition, LCSSP). Allocator E abandoned 6 Chinese yuan silently, and the participant learned about it accidently (the high-cost silent self-punishment condition, HCSSP). Specifically, in the communicative self-punishment conditions, the participant imaged that after the allocation, the allocator came to the participant's room and said "I decide to abandon 2/6 Chinese yuan and I have left it on the table in my room." In the silent self-punishment conditions, the participant imaged that after the allocation he or she accidently saw the allocator leave 2/6 Chinese yuan on the table when passing by the allocator's room. The allocator did not notice that he or she was being watched. The allocator did not tell anyone about his or her behavior till he or she left.

Participants were informed that the abandoned money would be taken back by the experimenter and would not be given to them. Due to this operation, the wrongdoer's abandoning money was not considered as a form of compensation. It is rather conceivable as an act of punishment against oneself (e.g., unilaterally incurring a certain amount of financial loss).

The participants were asked to what extent they forgave each allocator (1 to 7, 1 = *completely forgave*, 7 = *completely did not forgive*) and to what extent they thought doing nothing/abandoning 2 Chinese yuan/abandoning 6 Chinese yuan was costly (1 to 7, 1 = *not costly*, 7 = *very costly*).

Results

Ten participants were excluded due to misunderstanding the instructions, leaving 68 participants (41 females, $M_{age} = 22.4$ years, $SD_{age} = 2.6$ years) in the subsequent analysis.

We manipulated the cost of behaviors by varying the amount of money abandoned by the allocators. This manipulation was successful. When the cost ratings were submitted to a one-way analysis of variance (ANOVA), the main effect of the cost manipulation (doing nothing vs. abandoning 2 Chinese yuan vs. abandoning 6 Chinese yuan) was significant, $F(2, 134) = 107.17$, $p < .001$, partial $\eta^2 = .615$. Post-hoc analyses indicated that the participants thought abandoning 2 Chinese yuan was costlier ($M = 3.85$, $SD = 1.28$) than do nothing ($M = 2.15$, $SD = 1.57$, $p < .001$); abandoning 6 Chinese yuan ($M = 5.87$, $SD = 1.45$) was costlier than abandoning 2 Chinese yuan ($p < .001$) and then doing nothing ($p < .001$).

We compared participants' forgiveness ratings in the experimental conditions (LCCSP/HCCSP/LCSSP/HCSSP) with their forgiveness ratings in the control condition (doing nothing). The forgiveness ratings in all experimental conditions were significantly greater than the ratings in the control condition (all $ts > 8.41$, all $ps < .001$; Table 1).

A 2 (type: communicative self-punishment vs. silent self-punishment) × 2 (cost: high vs. low) repeated-measures ANOVA on forgiveness ratings was

Table 1. Means (and standard deviations) of forgiveness ratings in Study 1.

		Low cost		High cost	
	Control	CSP	SSP	CSP	SSP
Forgiveness	2.68 (1.41)	4.46 (1.64)	4.63 (1.62)	5.85 (1.52)	5.31 (1.71)

CSP: communicative self-punishment; SSP: silent self-punishment.

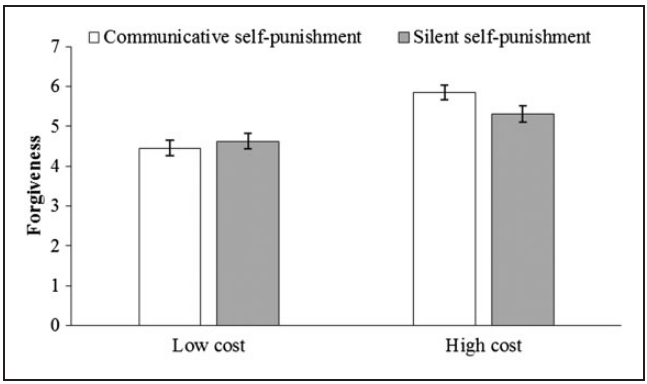

Figure 1. Mean forgiveness ratings (±SE) in different conditions in Study 1.

conducted. The main effect of type was not significant, $F(1, 67) = .84$, $p = .362$, partial $\eta^2 = .012$ (Figure 1). The main effect of cost ($F(1, 67) = 45.49$, $p < .001$, partial $\eta^2 = .404$) and the interaction effect ($F(1, 67) = 16.02$, $p < .001$, partial $\eta^2 = .193$) were significant. A simple effects analysis indicated that the effect of the type of self-punishment was significant in the high-cost level, $F(1, 67) = 5.17$, $p = .026$, but not in the low-cost level, $F(1, 67) = .08$, $p = .375$.

Discussion

The forgiveness ratings of the experimental conditions were all significantly greater than the ratings of the control condition. This result indicates that the wrongdoer's self-punishment does promote forgiveness. Consistent with costly signaling theory, the main effect of cost was significant. This result confirms our prediction that participants are more likely to forgive the wrongdoer who incurred a greater cost to punish themselves. An interaction effect between type and cost was found. At the high-cost level, participants were more inclined to forgive the communicative self-punisher than the silent self-punisher. However, at the low-cost level, the effect was not obtained. We would discuss this interaction effect in Study 2.

One limitation of Study 1 is that the within-subject design is used. Another limitation is that self-punishment is not compared with other methods (e.g., apology) which could promote forgiveness. To provide more convincing evidence, a between-subject design would be used and an apology condition would be added in Study 2.

Study 2

Method

Participants, design and procedure. A total of 287 undergraduate students participated in the study for course credits. They finished the experiment after class in the classroom. The design and the procedure of Study 2 were the same as Study 1 except for the following parts: (1) Study 2 had a between-subject design. (2) A new condition that the allocator said "sorry" to the participant after the allocation (the apology condition, AP) was added. (3) Participants were asked to what extent they thought the allocator's behavior after the allocation was costly.

Results

Thirty participants were excluded due to misunderstanding the instructions or failing to finish all the tasks, leaving 257 participants (119 females, $M_{age} = 18.37$ years, $SD_{age} = 1.15$ years) in the subsequent analysis.

The cost ratings in the experimental conditions (AP/LCCSP/HCCSP/LCSSP/HCSSP) were significantly or marginally greater than the ratings in the control condition, all ts > 1.92, all ps $< .058$ (Table 2). The cost ratings in the HCCSP/HCSSP conditions were significantly greater than the ratings in the AP condition, all ts > 5.13, all ps $< .001$. No significant difference in the cost ratings was found between the LCCSP/LCSSP conditions and the AP condition, all ts $< .75$, all ps $> .450$. A 2 (type: communicative self-punishment, silent self-punishment)$\times$2 (cost: high, low) repeated-measures ANOVA on the cost ratings was conducted. Importantly, the main effect of cost was significant, $F(1, 168) = 64.19$, $p < .001$, partial $\eta^2 = .276$. The main effect of type and the interaction

Table 2. Means (and standard deviations) of cost ratings in Study 2.

			Low cost		High cost	
	Control	Apology	CSP	SSP	CSP	SSP
Perceived cost	2.10 (1.02)	2.77 (1.51)	2.86 (1.44)	2.56 (1.18)	4.85 (2.06)	4.72 (2.00)

CSP: communicative self-punishment, SSP: silent self-punishment.

Table 3. Means (and standard deviations) of forgiveness ratings in Study 2.

			Low cost		High cost	
	Control	Apology	CSP	SSP	CSP	SSP
Forgiveness	4.32 (1.49)	5.07 (1.52)	4.80 (1.89)	5.04 (1.74)	6.00 (1.13)	5.23 (1.65)

CSP: communicative self-punishment, SSP: silent self-punishment.

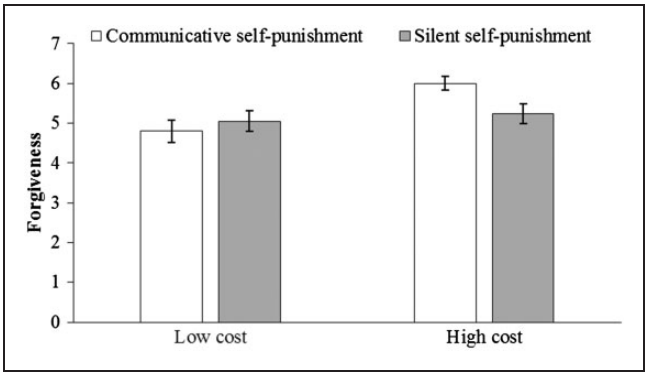

Figure 2. Mean forgiveness ratings (±SE) in different conditions in Study 2.

effect were not significant, $F(1, 168) = .71$, $p = .400$, partial $\eta^2 = .004$, $F(1, 168) = .12$, $p = .730$, partial $\eta^2 = .001$.

The forgiveness ratings in the AP/HCCSP/LCSSP/HCSSP conditions were significantly greater than the ratings in the control condition, all $ts > 2.07$, all $ps < .042$ (Table 3). There was no significant difference in forgiveness ratings between the LCCSP condition and the control condition, $t(83) = 1.13$, $p = .200$. The forgiveness ratings in the HCCSP conditions were significantly greater than the ratings in the AP condition, $t(82) = 3.17$, $p = .002$. No significant difference was found between the LCCSP/LCSSP/HCSSP conditions and the AP condition, all $ts < .75$, all $ps > .457$.

A 2 (type: communicative self-punishment vs. silent self-punishment) × 2 (cost: high vs. low) repeated-measures ANOVA on forgiveness ratings was conducted. The main effect of type was not significant, $F(1, 168) = 1.09$, $p = .297$, partial $\eta^2 = .006$ (Figure 2). The main effect of cost ($F(1, 168) = 7.90$, $p = .006$, partial $\eta^2 = .045$) and the interaction effect ($F(1, 168) = 4.21$, $p = .042$, partial $\eta^2 = .024$) were significant. A simple effects analysis showed that the effect of the type was significant in the high-cost level, $F(1, 169) = 4.16$, $p = .043$, but not in the low-cost level, $F(1, 169) = .47$, $p = .493$.

Discussion

In Study 2, the forgiveness ratings were greater in the AP than control condition. Replicating previous findings, compared with doing nothing, apology makes the victim be more inclined to forgive the wrongdoer (McCullough et al., 2013). The forgiveness ratings in the self-punishment conditions (expect for the LCCSP condition) were greater than them in the control condition. Consist with the results of Study 1, it indicates that self-punishment promotes forgiveness. Moreover, it was found the forgiveness ratings were greater in the HCCSP than AP condition, which means the effect of self-punishment could not only better than doing nothing, but also better than apology in certain condition.

Consistent with costly signaling theory, the significant main effect of cost was also found in Study 2. This result confirms our prediction that participants were more willing to forgive the wrongdoer who punished themselves more severely.

The interesting interaction effect between type and cost was found again. The effect of type on forgiveness was only significant at the high-cost level. There could be various explanations. One of the potential reasons is that when the self-punisher incurs a high cost to punish themselves, people think the self-punisher is truly remorseful rather than pretending to be regretful for their own self-interest. In this case, people hope that the self-punisher informs others about his or her self-punishment to avoid misunderstanding because without such information people may think the silent self-punisher does nothing after the unfair allocation and even incur cost to inflict unnecessary punishment on the wrongdoer. Thus, in this study, when the cost of self-punishment was high, participants tended to blame the silent self-punisher for not telling them about the self-punishment. When the self-punisher incurred a low cost to punish themselves, participants were not sure about the intention of the self-punisher and did not care about whether the self-punisher informed them about the self-punishment.

In Studies 1 and 2, we had studied how self-punishment affected forgiveness in the direct reciprocity context. In Study 3, we are going to preliminarily explore how self-punishment affects forgiveness in the indirect reciprocity context.

Study 3

Method

Participants, design and procedure. A total of 78 undergraduate students participated in the study for a monetary reward. The design and procedure of Study 3 were the same as Study 1 except for the following parts: (1) participants were asked to imagine that they were bystanders and watched five pairs of allocators and receivers playing the allocation game. Their task was to judge the behavior of different allocators. The different allocators' allocations and their consequent

Table 4. Means (and standard deviations) of forgiveness ratings in Study 3.

		Low cost		High cost	
	Control	CSP	SSP	CSP	SSP
Forgiveness	2.89 (1.64)	4.46 (1.64)	4.60 (1.51)	5.36 (1.77)	5.41 (1.81)

CSP: communicative self-punishment, SSP: silent self-punishment.

behaviors toward the receivers were the same as in Study 1. (2) The participants were required to answer questions from the standpoint of a bystander rather than the receiver.

Results

Eight participants were excluded due to misunderstanding the instructions, leaving 70 participants (41 females, $M_{age} = 22.4$ years, $SD_{age} = 2.6$ years) in the subsequent analysis.

A one-way ANOVA revealed that participants thought the cost of doing nothing/abandoning 2 Chinese yuan/abandoning 6 Chinese yuan varied significantly, $F(2, 134) = 200.43$, $p < .001$, partial $\eta^2 = .744$. Post-hoc analyses indicated that participants thought abandoning 2 Chinese yuan was costlier ($M = 3.56$, $SD = 1.25$) than doing nothing ($M = 1.73$, $SD = 1.06$, $p < .001$); abandoning 6 Chinese yuan ($M = 5.77$, $SD = 1.40$) was costlier than abandoning 2 Chinese yuan ($p < .001$) and then doing nothing ($p < .001$).

The forgiveness ratings in the experimental conditions were significantly greater than the ratings in the control condition (all $ts > 5.72$, all $ps < .001$; Table 4).

A 2 (type: communicative self-punishment, silent self-punishment) $\times$ 2 (cost: high, low) repeated-measures ANOVA on the forgiveness ratings was conducted. The main effect of cost was significant, $F(1, 69) = 28.72$, $p < .001$, partial $\eta^2 = .294$ (Figure 3). The main effect of type and the interaction effect were not significant, $F(1, 69) = .34$, $p = .562$, partial $\eta^2 = .005$, $F(1, 69) = .241$, $p = .625$, partial $\eta^2 = .003$.

Discussion

Consistent with Studies 1 and 2, the forgiveness ratings of the experimental conditions were all significantly greater than the ratings of the control condition in Study 3. This result indicates that the self-punishment also promotes forgiveness in indirect reciprocity. In accordance with costly signaling theory, the main effect of cost was significant, which implies the more cost the wrongdoer incurs, the more likely the bystander would forgive him.

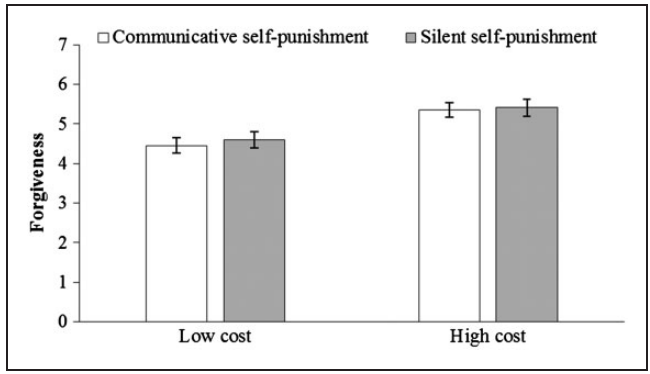

Figure 3. Mean forgiveness ratings (±SE) in different conditions in Study 3.

However, the interaction effect between type and cost on forgiveness was not significant in Study 3. It seemed that as bystanders, the participants in Study 3 were less sensitive to the wrongdoer's behavior than the participants in Studies 1 and 2, who played the role of victims. Schilbach et al. (2013) proposed that people behave differently when interacting with others than observing them. Feelings of involvement influence participants' thoughts and decisions. When the participants were victims, they were more involved in the economically harmful event because their self-interest was exploited. As victims, participants paid significant attention to the wrongdoers' consequent behavior and judged the cost and type of self-punishment carefully in case the exploitation happened again. By definition, bystanders are not directly involved in the wrongdoing. Therefore, participants in Study 3 might have paid less attention to the type of self-punishment than those in Studies 1 and 2.

General discussion

The present studies investigated the effect of self-punishment on forgiveness. In Studies 1, 2 and 3, the forgiveness ratings of the experimental conditions were significantly greater than the ratings of the control condition. These results show that in both the direct and indirect reciprocity contexts, self-punishment does promote forgiveness.

Confirming the prediction of costly signaling theory (Gintis et al., 2001; Roberts, 1998; Sosis, 2000; Zahavi, 1977), when the cost of self-punishment increases, the effect of self-punishment on forgiveness becomes stronger. The main effect of cost existed in both the direct and indirect reciprocity contexts. This result suggests that if someone wants to gain forgiveness from others, he or she should not spare the cost of self-punishment.

We also identified an interaction between the cost and type of self-punishment on forgiveness in Studies 1 and 2. In the direct reciprocity case, the participants were more inclined to forgive the communicative self-punisher than the silent self-punisher only at the high-cost level. A potential explanation is provided. People may consider the wrongdoer who engages in high-cost self-punishment as a good potential cooperator and hopes to keep a positive relationship with him. People may blame the self-punisher who engages in high-cost self-punishment silently because this behavior may result in misunderstanding. In contrast, people are not sure whether the low-cost self-punisher is a truly good cooperator; therefore, they do not focus on whether he or she informed others about his or her self-punishment. These results imply that to obtain forgiveness, people should inform victims about their self-punishment, especially when the cost of the self-punishment is high.

The interaction between the cost and type of self-punishment was not observed in Study 3. It could be caused by fewer feelings of involvement in the indirect reciprocity than direct reciprocity context (Schilbach et al., 2013). In the direct reciprocity studies, because participants' self-interest was directly harmed by the wrongdoer in the imaginary scenery, they would carefully decide whether to forgive the wrongdoer according to both the cost and type of self-punishment. In contrast, in the indirect reciprocity study, because the participants only observed and their self-interest was not harmed, they roughly formed their attitude towards the wrongdoer only based on the cost of the self-punishment. These results imply that people judged the wrongdoer's remorseful behavior more carefully in the direct reciprocity case compared with indirect reciprocity.

In our research, we studied economic self-punishment. However, wrongdoers can use different methods to punish themselves in terms of money, time, energy or pain (Bastian et al., 2011; Inbar et al., 2013; Ohtsubo & Watanabe, 2009; Watanabe & Ohtsubo, 2012). An interesting research direction is to investigate whether different types of self-punishment have the same effect on forgiveness.

A limitation of our research is that we only use a within-subject design (Study 3) to preliminary explore the self-punishment's effect on forgiveness in the indirect reciprocity context. Future studies may test whether the effect of self-punishment in the indirect reciprocity context still exists when using a between-subject design. Besides, though a possible reason is provided, the present studies could not directly explain the interaction effect between cost and type on forgiveness in the direct reciprocity context. More studies are needed to illustrate this interesting phenomenon.

In conclusion, the present studies demonstrate that self-punishment promotes forgiveness in both direct and indirect reciprocity contexts. The costlier the self-punishment is, the stronger the effect it has on forgiveness. In the direct reciprocity context, people are more likely to forgive the communicative

self-punisher than the silent self-punisher when the cost of self-punishment is high. These findings can guide us regarding how to address a damaged relationship via self-punishment when compensation is not feasible or acceptable.

Acknowledgments

The authors thank two anonymous reviewers for constructive and helpful comments.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Basic Research Program of China (2011CB711000, 2013CB837300), the National Natural Science Foundation of China (NSFC) (31170971, 61210010,31400888), the Major Project of National Social Science Foundation (12&ZD228), the Beijing Municipal Science & Technology Commission (Z151100003915122), and the National Program for Support of Top-notch Young Professionals.

References

Bastian, B., Jetten, J., & Fasoli, F. (2011). Cleansing the soul by hurting the flesh: The guilt-reducing effect of pain. *Psychological Science*, *22*(3), 334–335.

Baumard, N., André, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, *36*(1), 59–78.

Berman, J. Z., Levine, E. E., Barasch, A., & Small, D. A. (2015). The Braggart's Dilemma: On the social rewards and penalties of advertising prosocial behavior. *Journal of Marketing Research*, *52*(1), 90–104.

Cashdan, E. A. (1980). Egalitarianism among hunters and gatherers. *American Anthropologist*, *82*(1), 116–120.

Eaton, J., Struthers, C. W., & Santelli, A. G. (2006). The mediating role of perceptual validation in the repentance–forgiveness process. *Personality and Social Psychology Bulletin*, *32*(10), 1389–1401.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87.

Gauché, M., & Mullet, E. (2005). Do we forgive physical aggression in the same way that we forgive psychological aggression? *Aggressive Behavior*, *31*(6), 559–570.

Gintis, H., Smith, E., & Bowles, S. (2001). Costly signaling and cooperation. *Journal of Theoretical Biology*, *213*(1), 103–119.

Inbar, Y., Pizarro, D. A., Gilovich, T., & Ariely, D. (2013). Moral masochism: On the connection between guilt and self-punishment. *Emotion*, *13*(1), 14–18.

Leff, G. (1999). *Heresy in the later middle ages: The relation of heterodoxy to dissent*. Manchester, UK: Manchester University Press.

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2013). Cognitive systems for revenge and forgiveness. *Behavioral and Brain Sciences*, *36*(01), 1–58.

Nelissen, R. M. A. (2012). Guilt-induced self-punishment as a sign of remorse. *Social Psychological and Personality Science*, *3*(2), 139–144.

Nelissen, R. M. A., & Zeelenberg, M. (2009). When guilt evokes self-punishment: Evidence for the existence of a Dobby Effect. *Emotion*, *9*(1), 118–122.

Ohtsubo, Y., Matsunaga, M., Komiya, A., Tanaka, H., Mifune, N., & Yagi, A. (2014). Oxytocin receptor gene (OXTR) polymorphism and self-punishment after an unintentional transgression. *Personality and Individual Differences*, *69*, 182–186.

Ohtsubo, Y., & Watanabe, E. (2009). Do sincere apologies need to be costly? Test of a costly signaling model of apology. *Evolution and Human Behavior*, *30*(2), 114–123.

Ohtsubo, Y., Watanabe, E., Kim, J., Kulas, J. T., Muluk, H., Nazar, G., . . . Zhang, J. (2012). Are costly apologies universally perceived as being sincere? *Journal of Evolutionary Psychology*, *10*(4), 187–204.

Roberts, G. (1998). Competitive altruism: From reciprocity to the handicap principle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1394), 427–431.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., . . . Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414.

Smith, E. A., & Bird, R. L. B. (2000). Turtle hunting and tombstone opening: Public generosity as costly signaling. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society*, *21*(4), 245–261.

Sosis, R. (2000). Costly signaling and torch fishing on Ifaluk atoll. *Evolution and Human Behavior*, *21*(4), 223–244.

Tanaka, H., Ohtsuki, H., & Ohtsubo, Y. (2016). The price of being seen to be just: An intention signalling strategy for indirect reciprocity. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1835), 20160694.

Tanaka, H., Yagi, A., Komiya, A., Mifune, N., & Ohtsubo, Y. (2015). Shame-prone people are more likely to punish themselves: A test of the reputation-maintenance explanation for self-punishment. *Evolutionary Behavioral Sciences*, *9*(1), 1–7.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.

Wallington, S. A. (1973). Consequences of transgression: self-punishment and depression. *Journal of Personality and Social Psychology*, *28*(1), 1–7.

Watanabe, E., & Ohtsubo, Y. (2012). Costly apology and self-punishment after an unintentional transgression. *Journal of Evolutionary Psychology*, *10*(3), 87–105.

Zahavi, A. (1977). The cost of honesty: further remarks on the handicap principle. *Journal of Theoretical Biology*, *67*(3), 603–605.

Author Biographies

Ruida Zhu is a PhD student in State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University. His research focuses on social emotion and social behavior.

Xueyi Shen is a master student in State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University. Her research focuses on social decision making and depression.

Honghong Tang is a PhD student in State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University. Her research focuses on social decision making and moral behavior.

Peixia Ye is a master student in State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University. Her research focuses on social decision making.

Huagen Wang is a master student in State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University. Her research focuses on social decision making and stress.

Xiaoqin Mai is an associate professor in Department of Psychology, Renmin University of China. Her research focuses on social decision making.

Chao Liu is a professor in State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University. His research focuses on social decision making and moral behavior.