Original Article

Relational Utility Affects Self-Punishment in Direct and Indirect Reciprocity Situations

Ruida Zhu,[1,2] Tao Jin,[1,2] Xueyi Shen,[1,2] Shen Zhang,[1,2] Xiaoqin Mai,[3] and Chao Liu[1,2]

[1]State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, PR China
[2]Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, PR China
[3]Department of Psychology, Renmin University of China, Beijing, PR China

Abstract: Previous studies of self-punishment focused on negative emotions and information transmission between wrongdoers and victims. We propose that self-punishment can be moderated by relational utility and can work not only in direct but also indirect reciprocity. In Studies 1 and 2, participants were more inclined to punish themselves when the victim could benefit the participants in future interactions than when the victim could not. In Study 3, participants were more inclined to punish themselves when the bystander could potentially offer lots of benefits to them in the future compared to when the bystander could only offer few or no benefits. These findings support our hypothesis, suggesting that wrongdoers strategically use self-punishment to pursue profits through repairing damaged relationships which are really conducive to achieve their personal goals. It helps us to understand self-punishment better in real life.

Keywords: self-punishment, relational utility, direct reciprocity, indirect reciprocity

In human society, wrongdoers sometimes exert economic loss or physical damage to themselves after violating social norms (Bastian, Jetten, & Fasoli, 2011; Inbar, Pizarro, Gilovich, & Ariely, 2013; Nelissen, 2011; Nelissen & Zeelenberg, 2009; Tanaka, Yagi, Komiya, Mifune, & Ohtsubo, 2015; Watanabe & Ohtsubo, 2012). Such a phenomenon, referred to as self-punishment, attracts researchers' interest as it seems to diminish wrongdoers' own benefit and does not benefit anyone.

Researchers have proposed two possible explanations. First, self-punishment may be driven by negative emotions, especially guilt, due to harming others (Nelissen, 2011; Nelissen & Zeelenberg, 2009; Tanaka et al., 2015; Watanabe & Ohtsubo, 2012). Studies have found that the guiltier the participants felt, the more severe punishment they inflicted on themselves; in turn, the more severe punishment the participants gave themselves, the more their guilt was relieved (Bastian et al., 2011; Inbar et al., 2013). Second, self-punishment may be a method for wrongdoers to express their remorse to their victims. If the option of direct compensation is unavailable, wrongdoers can choose to engage in self-punishment to express their remorse after committing a transgression, which is conductive to relationship maintenance (Nelissen & Zeelenberg, 2009). Nelissen (2011) further found that wrongdoers are more willing to engage in self-punishment in the presence of a victim than in the presence of a general audience. It implies self-punishment is used as a signal of remorse for victims specifically.

Existing studies suggest that self-punishment is driven by the guilt of violating moral standards and is used to express remorse (Bastian et al., 2011; Nelissen, 2011; Nelissen & Zeelenberg, 2009; Watanabe & Ohtsubo, 2012), which implies that self-punishment is mainly motivated by moral and other-focused considerations (e.g., victims are harmed). However, self-punishment may also be motivated by some self-interested and self-focused considerations (e.g., the potential benefits that self-punishment may bring to wrongdoers).

It is generally assumed that moral behaviors benefit people in the long run by maintaining reciprocal rewarding relationships (Baumard, André, & Sperber, 2013; Haidt, 2003; Trivers, 1971). However, this hypothesis is correct only if the long-term benefits of the moral behavior surpass its costs. Thus, it is necessary for people to adjust their responses according to the potential benefits. Supporting this hypothesis, recently Nelissen (2014) found that moral emotional guilt, which is closely related to moral behavior, is moderated by relational utility, the utility of others for the achievement of one's personal aims through social

© 2017 Hogrefe Publishing

DOI: 10.1027/1864-9335/a000291

interaction. This finding was replicated in Ohtsubo and Yagi (2015)'s research, which showed that the wrongdoers' feelings of guilt following interpersonal transgressions increased with an increase of the victims' relational utility. Because guilt serves as an emotional mechanism for protecting interpersonal relationships (Haidt, 2003), one may infer that with regard to long-term benefits, wrongdoers are more likely to repair a damaged relationship that could bring them more benefits. A wrongdoer's guilt per se could not repair a relationship, while guilt-related moral behaviors such as self-punishment that signal remorse, could (Nelissen, 2011; Ohtsubo & Watanabe, 2009). Therefore, it is probable that relational utility affects guilt-related behavior such as self-punishment as well. So our first aim in this study is to test whether the effect of relational utility on the feelings of guilt could extend to self-punishment.

Most previous research on relational utility focused on the interaction between wrongdoers and victims, and thus only studied the effect of direct reciprocity (Nelissen, 2014; Ohtsubo & Yagi, 2015). However, according to the indirect reciprocity theory, wrongdoers tend to display their benign intentions through specific behaviors to maintain a positive reputation in public, as bystanders also judge wrongdoers' behavior, and many of them are willing to bear some costs to punish wrongdoers (Nowak & Sigmund, 1998; Fehr & Fischbacher, 2004). So wrongdoers may also take the relational utility of bystanders into consideration when making a decision about self-punishment in front of them. Our second aim in this study thus is to test whether relational utility influences wrongdoers' self-punishment in both direct and indirect reciprocity contexts.

Overview of the Current Research

In the present research, relational utility refers to the potential monetary benefits the victim or bystander could offer to the wrongdoer in the future. Self-punishment is defined as participants' behavior of abandoning their own monetary benefits. In Studies 1 and 2, we tested whether relational utility affects self-punishment in direct reciprocity by manipulating participants' future chance of reciprocating with the victim. The future chance of reciprocity increases the utility of a relationship as people could benefit from future cooperation (Trivers, 1971). In Study 3, we tested whether relational utility affects self-punishment in indirect reciprocity. Besides the future opportunity for reciprocity, we also manipulated the amount of benefits the bystander was able to offer. The relational utility increases with increases in the amount of benefits that the bystander could provide (Baumard et al., 2013). We predict that relational utility promotes self-punishment in both direct and indirect reciprocity situations. All studies were approved by the Institutional Review Board of Beijing Normal University.

Study 1

Methods

Participants and Design

Forty-three undergraduate students (23 females, M_{age} = 22.1 years, SD_{age} = 2.1 years) participated in the study for payment. The study had a one-factor (Future opportunity for reciprocity: Future vs. No future), between-subject design.

Procedure

The participants came in groups and performed tasks on the computer alone in separate rooms. The participants in each group did not know each other before the experiment.

Personality Measurements

The Interpersonal Reactivity Index (IRI) scale (Davis, 1980), which measures participants' dispositional capacity for empathy, was completed by the participants. Tangney and Dearing (2003) found that people's dispositional capacity for empathy is positively correlated to their feelings of guilt. As guilt is closely related to self-punishment, it is necessary to measure people's capacity for empathy.

Damage to Relationship

The participants who were assigned role A played three rounds of a time-estimation game, ostensibly with another player (B) (adapted from Nelissen & Zeelenberg, 2009, Study 2). In each round, the participants and B played 10 trials of the game independently to earn points for themselves or for each other. They were told that the more points the participants owned (the points that the participants earned for themselves and the points that the other player earned for them), the more monetary rewards they would receive. In each trial, a red lamp appeared for 2,000 ms, turned green, and remained lit until a corresponding key was pressed. The participants were asked to press the key when they estimated that the green lamp had been lit for 3,000 ms. Any estimation between 2,700 ms and 3,300 ms was considered correct. A correct estimation earned 10 points. Before the formal experiment, a practice round was provided to help participants familiarize themselves with the game.

During the formal experiment, at the beginning of the first round, the participants were informed that they would earn points for themselves. Regardless of their real performance, the predetermined overall feedback – that both the participants and B earned 80 points for themselves – was shown after the 10 trails of the first round were finished.

The feedback served as a reference for the participants' performance in the second round.

At the beginning of the second round, the participants were informed that they would earn points for each other. The predetermined feedback showed that B earned the participant 80 points, while the participant only earned B 30 points. According to the performance in the first round, the feedback from this round implied that the participants had damaged B's benefits, and according to the results, it seemed that the participants cared more about their own benefits than B's benefits. In such a situation, the participants' cooperative relationship with B could be damaged.

Future Opportunity for Reciprocity

At the beginning of the third round, the participants were informed that they would earn points for each other (the Future condition) or for themselves (the No future condition) and that this was the final round.

Self-Punishment

After learning of the recipient of the points in the third round, the participants were informed that all players could decide whether to deduct their own points, which ranged from 0 to all the points they had at that time. They were told that those deducted points just disappeared and would NOT be given to B, but a message about how many points they deducted would be sent to B (e.g., "A deducted himself or herself 5 points"). If they deducted zero points, no message would be sent. After participants deducted their points, they would find that B did not deduct any points. Afterwards, the third round was played without feedback. At the end, the participants were told that the feedback would be shown after they finished the other measurements.

Emotion Measurements

When the time-estimation task of the third round was finished, the participants were asked to rate (1 = *very slightly or not at all*, 5 = *extremely*) how guilty and how distressed and upset (two guilt-like emotions) they felt when they saw the feedback of the second round.

Debriefing

The participants were probed using a funnelled procedure that tested the participants' comprehension of the instructions, general suspicions about the authenticity of the feedback, and interaction.

Results

Three participants were excluded due to not understanding the experimental instructions or having suspicions about the authenticity of the feedback, leaving 40 (20 in the Future condition and 20 in the No future condition) in the subsequent analyses.

Guilt

To test whether Nelissen's (2014) finding that relational utility affects guilt could be replicated, an independent-samples *t*-test on guilt ratings was run. The feeling of guilt was not significantly different between Future (M = 3.60, SD = 1.31) and No future (M = 3.55, SD = 1.50) conditions, $t(38) = .11$, $p = .911$, Cohen's $d = .04$.

Self-Punishment

To test whether future reciprocity affects self-punishment, a one-way (Future opportunity for reciprocity) ANOVA on the number of self-deducted points was run. Consistent with our hypothesis that relational utility promotes self-punishment, the number of self-deducted points in the Future condition (M = 42.70, SD = 27.97) was significantly higher than that in the No future condition (M = 20.75, SD = 22.14), $F(1, 38) = 7.57$, $p = .009$, $\eta^2 = .070$.

As studies have found that empathy is positively correlated to guilty feelings (Tangney & Dearing, 2003) and that negative feelings are positively correlated to self-punishment (Inbar et al., 2013), we ran an analysis of covariance (ANCOVA) to test whether relational utility could affect self-punishment independent of these emotion-related factors. After controlling for the ratings of guilt, distress, upset, and empathy, the ANCOVA revealed that the difference in self-deducted points between the Future and No future conditions was still significant: $F(1, 34) = 6.33$, $p = .017$, $\eta^2 = .065$. There were no significant effects from these covariates on self-deducted points (all *F*s < 1.28, *ns*).

To test whether wrongdoers punish themselves when no relational utility is involved, a one-sample *t*-test was run on the number of self-deducted points in the No future condition. The number of self-deducted points in the No future condition was significantly higher than 0, $t(19) = 4.19$, $p < .001$.

Discussion

Confirming our prediction, the participants were significantly more inclined to engage in self-punishment in the Future than No future condition, even when the effects of negative emotions and empathy were controlled. Because the future chance for reciprocity increases the utility of a relationship, these results suggest that relational utility promotes self-punishment. Additionally, in the No future condition, the number of self-deducted points was significantly higher than 0, which indicated that the participants engaged in self-punishment when the victim could not offer any benefits in the future. The results are consistent with previous research that found that self-punishment is (partly) driven by guilt due to transgression (Bastian et al., 2011; Inbar et al., 2013).

In Study 1, some methodological limitations may be present. First, emotional feelings were measured after the participants made their self-punishment decisions, which meant the participants' feelings might have been regulated by the self-punishment. That may be why we did not replicate Nelissen's (2014) finding that rational utility affects feelings of guilt. Second, we did not have control conditions in which the reciprocal relationship was not damaged. Third, we did not directly ask the participants to judge their reciprocal relationship with their partner, so our hypothesis that the participants could realize the damage to their reciprocal relationship was not supported directly. We would attempt to resolve these problems in Study 2.

Study 2

Methods

Participants and Design

A total of 138 undergraduate students (106 females, M_{age} = 21.6, SD_{age} = 2.3) participated in the study for payment.[1] The study had a 2 (Relationship status: Damaged vs. Non-damaged) × 2 (Future opportunity for reciprocity: Future vs. No future) between-subject design.

Procedure

The basic rules and procedures of Study 2 were exactly the same as those of Study 1 except for the following elements: (1) The personality measure was finished at least two days before the experiment. (2) A new factor (Relationship status: Damaged vs. Non-damaged) was added. In the two new control conditions (Non-damaged-Future and Non-damaged-No future conditions), the participants and B earned themselves 80 points in the first round and earned each other 80 points in the second round. In such cases, the reciprocal relationships were not damaged. The other two conditions (Damaged-Future and Damaged-No future conditions) were the same as the two conditions (Future and No future conditions) in Study 1. (3) The participants' emotions were measured after the participants knew for whom they would earn points in the third round and before their self-punishment decision. Here, we also asked the participants to rate their reciprocal relationship with B at the moment (1 = *very negative*, 7 = *very positive*). (4) When the game was finished, self-punishers in the Damaged-Future and Non-damaged-Future conditions answered two questions in order to examine whether they knew self-punishment might benefit them: (a) After B knew how many points you deducted, how many points did you expect to receive from B in the third round? and (b) Assuming that you did not deduct any points, how many points would you expect to receive from B in the third round? The participants who did not deduct any points from themselves did not need to answer these questions.

Results

Twelve participants were excluded due to misunderstanding the experimental instructions or having suspicions regarding the authenticity of the feedback, leaving 126 (33 in the Damaged-Future condition, 32 in the Damaged-No future condition, 31 in the Non-damaged-Future condition, and 30 in the Non-damaged-No future condition) in the subsequent analyses.

Reciprocal Relationship

To check the manipulation of relationship status, a 2 (Relationship status) × 2 (Future opportunity for reciprocity) analysis of variance (ANOVA) on reciprocal relationship ratings was conducted. The main effect of the relationship status was significant, $F(1, 122) = 380.55$, $p < .001$, $\eta^2 = .100$ (Table 1), which meant participants could realize the damage to their reciprocal relationship. Neither the main effect of future opportunity for reciprocity, $F(1, 122) = .74$, $p = .393$, $\eta^2 < .001$, nor the interaction effect was significant $F(1, 122) = 1.19$, $p = .277$, $\eta^2 < .001$.

Guilt

To examine whether relational utility affects guilt, a 2 (Relationship status) × 2 (Future opportunity for reciprocity) ANOVA on guilt ratings was run. It revealed that the main effect of the relationship status was significant, $F(1, 122) = 285.96$, $p < .001$, $\eta^2 = .198$, which implied the participants who harmed the other's economic benefits were more guilty than the participants who did not. The main effect of the future opportunity for reciprocity was not significant, $F(1, 122) = 2.11$, $p = .15$, $\eta^2 = .001$. Importantly, the interaction effect was significant, $F(1, 122) = 7.54$, $p = .007$, $\eta^2 = .005$. A simple effects

[1] In the beginning, we recruited 88 and 120 participants in Studies 2 and 3, respectively. As the sample size was small, we decided to run an additional population of participants to ensure the reliability of our results. Using G*Power 3 software (Faul, Erdfelder, Lang, & Buchner, 2007), we determined that the minimum sample size was 29 participants per condition, which could provide adequate power ($1 - \beta > .80$) and medium-sized effect (f = .25). To meet this standard, we ran additional 50 participants in Study 2 and additional 6 participants in Study 3. The additional samples did not change the statistical results of Studies 2 and 3, so we only report the results when the additional samples were included.

Table 1. Means (and standard deviations) of negative feelings, empathy, reciprocal relationship ratings, and self-deducted points in Study 2

	Damaged		Non-damaged	
	Future	No future	Future	No future
Guilt	4.09 (1.04)	3.44 (1.24)	1.03 (0.18)	1.23 (0.57)
Distress	2.42 (0.90)	2.22 (1.26)	1.26 (0.58)	1.53 (0.78)
Upset	2.18 (1.18)	2.00 (1.11)	1.42 (0.85)	1.83 (1.05)
Empathy	3.43 (0.38)	3.33 (0.45)	3.56 (0.28)	3.49 (0.31)
Relationship	3.30 (1.40)	3.34 (1.00)	6.87 (0.34)	6.53 (0.78)
Deducted points	32.27 (16.82)	12.19 (19.30)	0.16 (0.90)	1.67 (5.92)

analysis showed that in the Damaged condition, participants felt marginally more guilt when the future opportunity for reciprocity was present compared to it was absent, $F(1, 123) = 3.08$, $p = .082$, $\eta^2 = .024$, but in the No-damaged condition they did not feel more guilt when the future opportunity for reciprocity was present compared to when it was absent, $F(1, 123) = .36$, $p = .547$, $\eta^2 = .003$. As the relational utility is manipulated by the future opportunity for reciprocity, these results suggest relational utility affects guilt when people do commit a transgression, which is consistent with Nelissen (2014)'s finding.

Self-Punishment

To examine whether relational utility affects self-punishment, a 2 (Relationship status) × 2 (Future opportunity for reciprocity) ANOVA on self-deducted points was run. There was a significant main effect of the future opportunity for reciprocity, $F(1, 122) = 15.31$, $p < .001$, $\eta^2 = .045$, and a significant main effect of the relationship status, $F(1, 122) = 80.58$, $p < .001$, $\eta^2 = .234$. Importantly, the interaction effect was also significant, $F(1, 122) = 20.67$, $p < .001$, $\eta^2 = .060$ (Figure 1). A simple effects analysis showed that in the Damaged condition, participants deducted themselves more points when the future opportunity for reciprocity was present compared to when it was absent, $F(1, 123) = 23.15$, $p < .001$, $\eta^2 = .158$, but in the No-damaged condition they did not deduct themselves more points when the future opportunity for reciprocity was present compared to when it was absent, $F(1, 123) = .18$, $p = .671$, $\eta^2 = .001$.

To test whether relational utility could affect self-punishment independently, a follow-up ANCOVA, controlling for ratings of guilt, distress, upset, and empathy, was run. It revealed that the main effect of the future opportunity for reciprocity and the interaction effect were still significant, $F(1, 118) = 13.03$, $p < .001$, $\eta^2 = .034$, $F(1, 118) = 13.09$, $p < .001$, $\eta^2 = .034$. The main effect of the relationship status was marginally significant, $F(1, 118) = 3.57$, $p = .061$, $\eta^2 = .009$. Guilt had a significant effect on the self-deducted points, $F(1, 118) = 12.48$, $p = .001$, $\eta^2 = .033$. Not surprisingly, the guiltier the wrongdoers feel, the more severe are the punishments they inflicted on themselves (Inbar et al., 2013). There were no significant effects of the other covariates on the self-deducted points (all *F*s < 1.28, *ns*). These results suggest relational utility, independent of emotional factors, could affect self-punishment when people do commit a transgression.

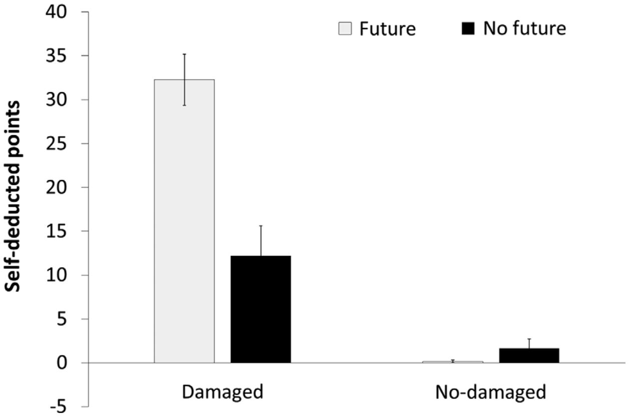

Figure 1. Mean self-deducted points (± *SE*) in different conditions in Study 2.

To test whether wrongdoers punish themselves when no relational utility is involved, a one-sample *t*-test was run on the number of self-deducted points in the Damaged-No future condition. The number of self-deducted points in the Damaged-No future condition was significantly higher than 0, $t(31) = 3.57$, $p = .001$.

Expectation

To test whether self-punishers know self-punishment might benefit them, a paired-samples *t*-test was run. In the Damaged-Future condition, the self-punishers (30 participants), who deducted themselves more than 0 point, expected that in the third round B would earn them more points after their self-punishment ($M = 69.67$, $SD = 12.72$) compared to if they had not punished themselves ($M = 56.67$, $SD = 21.38$), $t(30) = 3.29$, $p = .003$, Cohen's $d = .677$. These data in the Non-damaged-Future condition were not analyzed, as there was only one self-punisher.

Discussion

Consistent with Study 1, the participants punished themselves more severely for their transgressions if the victim was of higher relational utility. It is noteworthy that when negative emotions were measured before self-punishment, we replicated Nelissen's (2014) finding that relational utility intensified feelings of guilt following a transgression. After controlling the effects of negative emotions and empathy, the main effect of future opportunity for reciprocity and the interaction effect on self-punishment were still significant, which implies that relationship utility could directly affect self-punishment independent of negative emotions. Confirming that the participants realized that their reciprocal relationships with B were damaged when they earned B only 30 points, participants in the Damaged conditions, compared to the participants in the Non-damaged conditions, reported that their reciprocal relationships were more negative. In addition, the self-punishers in the Damaged-Future condition expected that B would earn them more points after they punished themselves than if they had not punished themselves. It indicates that in their minds, self-punishers know self-punishment may benefit them.

Together, Studies 1 and 2 suggest that relational utility can influence self-punishment independent of guilt in a direct reciprocity situation. Would this type of interest-oriented self-punishment be found in an indirect reciprocity situation? We would explore this question in Study 3.

Study 3

Methods

Participants and Design

A total of 126 undergraduate students (90 females, M_{age} = 22.0, SD_{age} = 2.2) participated in the study for payment. The study had a 2 (Amount of potential benefits: More vs. Less) × 2 (Future opportunity for reciprocity: Future vs. No future) between-subject design.

Procedure

The basic rules and procedures of the game in Study 3 were similar to those of Study 1. The difference was that the participant played the game ostensibly with two other players (B as the victim and C as the bystander).

Damage to Relationship

In the first round, all players earned points for themselves. The feedback showed that the participant and B earned 80 points for themselves. In the second round, the participant and B earned points for each other while C earned points for himself or herself. The feedback showed that B earned the participant 80 points, while the participant earned B only 30.

The Amount of Potential Benefits and Future Opportunity for Reciprocity

The feedback showed that in the first and second rounds, C earned himself or herself 70 and 90 points (the More condition) or 40 and 30 points (the Less condition).

At the beginning of the third round, the participants were informed that they and C would earn points for each other while B would earn points for himself (the Future condition) or that all players would earn points for themselves (the No future condition).

Self-Punishment

The participants were informed that the players could decide whether to deduct points from themselves. The participants were also told that randomly decided by a computer program a message about how many points they deducted would be sent to C (bystander) only, but NOT to B (victim).

Personality, Emotion Measurements, and Debriefing

The measurements and debriefing procedure were identical to those in Study 1.

Results

Ten participants were excluded from the following analysis due to either misunderstanding the experimental instructions or having suspicions about the authenticity of the interaction, leaving 116 participants (29 participants in each condition) in the subsequent analyses.

Guilt

To test whether relational utility affects guilt in an indirect reciprocity situation, a 2 (Future opportunity for reciprocity) × 2 (Amount of potential benefits) ANOVA on guilt ratings was conducted. There was no significant main effect or interaction effect (all *F*s < 1, *ns*) (Table 2).

Self-Punishment

To examine whether relational utility affects self-punishment in an indirect reciprocity situation, a 2 (Future opportunity for reciprocity) × 2 (Amount of potential benefits) ANOVA on self-deducted points was run. There was a significant main effect of the future opportunity for reciprocity, $F(1, 112) = 7.40$, $p = .008$, $\eta^2 = .037$, and a marginally significant main effect of the amount of potential benefits, $F(1, 112) = 3.75$, $p = .055$, $\eta^2 = .019$) (Figure 2). The interaction effect was not significant, $F(1, 112) = 1.41$, $p = .237$, $\eta^2 = .007$.

Table 2. Means (and Standard Deviations) of negative feelings, empathy, reciprocal relationship ratings, and self-deducted points in Study 3

	More		Less	
	Future	No future	Future	No future
Guilt	4.03 (0.94)	3.86 (1.33)	3.83 (1.44)	3.69 (1.20)
Distress	2.55 (1.24)	2.00 (1.04)	2.21 (1.26)	2.48 (1.43)
Upset	2.83 (1.44)	2.17 (1.36)	2.55 (1.15)	2.45 (1.43)
Empathy	3.37 (0.41)	3.44 (0.45)	3.55 (0.44)	3.36 (0.47)
Deducted points	20.52 (18.39)	9.52 (11.80)	11.72 (15.60)	7.41 (14.05)

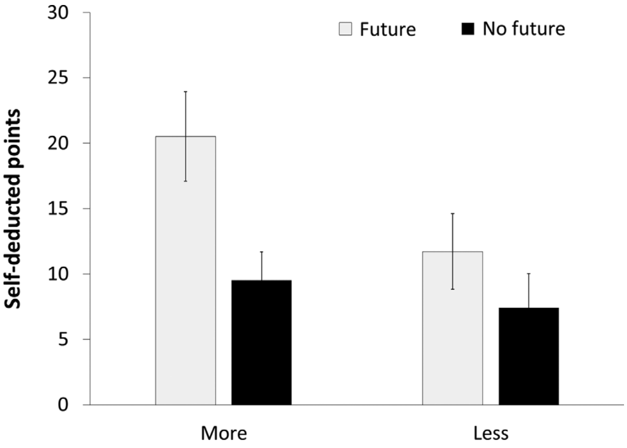

Figure 2. Mean self-deducted points (± *SE*) in different conditions in Study 3.

To test whether relational utility could affect self-punishment independently, a follow-up ANCOVA, controlling for ratings of guilt, distress, upset, and empathy, was conducted. It revealed that the main effect of the future opportunity for reciprocity was significant, $F(1, 108) = 7.18$, $p = .009$, $\eta^2 = .033$. The main effect of the amount of potential benefits was marginally significant $F(1, 108) = 3.04$, $p = .084$, $\eta^2 = .014$. The interaction effect was not significant, $F(1, 108) = .924$, $p = .339$, $\eta^2 = .004$. Guilt had a significant effect on the self-deducted points, $F(1, 108) = 6.40$, $p = .013$, $\eta^2 = .030$. There were no significant effects of the other covariates on the self-deducted points (all *F*s < 2.41, *ns*). As the bystander's relational utility is manipulated by the future opportunity for reciprocity and the amount of potential benefits, these results suggest relational utility, independent of emotional factors, could affect self-punishment in an indirect reciprocity situation.

To test whether wrongdoers punish themselves when no relational utility is involved in an indirect reciprocity situation, one-sample *t*-tests were run on the number of self-deducted points in the More-No future and the Less-No future conditions. The number of deducted points in the More-No future and the Less-No future conditions was significantly higher than 0, $t(28) = 4.34$, $p < .001$, $t(28) = 2.84$, $p = .008$.

Discussion

As the participants' self-punishment was shown only to the bystander, the significant effects of relational utility on self-punishment provide evidence that the participants took the relational utility of the bystander into consideration when making a decision about self-punishment. The results support our prediction that relational utility also exacerbates self-punishment in indirect reciprocal relationships. In addition, in the No future conditions, the participants' self-deducted points were significantly more than 0. This indicates that wrongdoers punish themselves even when the bystander could not provide any benefits for them, which confirms previous research findings that self-punishment is partly driven by guilt due to transgression (Inbar et al., 2013). In Study 3, feelings of guilt were not affected by the relational utility of the bystander. This is reasonable because the source of guilt is the harm inflicted on the victim. Guilty feelings are closely related to the victim rather than the bystander. Some researchers have proposed that shame instead of guilt may be affected by the relational utility of the bystander in an indirect reciprocity context (Nelissen, 2014). This needs to be confirmed in future studies.

General Discussion

The present research investigated the influence of relational utility on self-punishment. In Studies 1 and 2, the participants were more inclined to punish themselves when the victim could benefit them in the future compared to when the victim could not. In Study 3, the participants were more likely to engage in self-punishment when the bystander could potentially offer many benefits to them in the future compared to when the bystander could offer only a few or no benefits. These effects were still robust after controlling negative emotions and empathy. These results demonstrate that relational utility can directly influence self-punishment in both direct and indirect reciprocity situations. Thus we extend the effect of relational utility on feelings of guilt (Nelissen, 2014) to guilt-related behavior (self-punishment), which further supports the notion that moral behaviors bring long-term benefits to people through the maintenance of reciprocal rewarding relationships (Baumard et al., 2013).

Our results confirm the close relationship between self-punishment and guilt. After being induced to feel guilty, the participants punished themselves even when no benefits were involved. These results are consistent with previous findings that self-punishment is partly driven by guilt due to transgression (Bastian et al., 2011; Inbar et al., 2013). It indicates that the effects of emotions and cost-benefit analyses may be intermingled in a self-punishment decision.

Our findings that both the victim and the bystander are considered as potential receivers of self-punishment by participants appear to be inconsistent with the results of Nelissen's study (2011), in which participants seemed to use self-punishment as a signal of remorse for victims specifically. This difference is likely to be caused by different experimental manipulations. In our Study 3, the participants were clearly aware that their transgressions were known by the bystander. However, in the study by Nelissen (2011), the participants in the general audience condition were not sure whether the bystander knew about their transgression and thus did not inflict extra punishment on themselves.

There are some limitations of the present studies. First, there is not a good cover story for the possibility of deducting points from the participants' own private pool. Nevertheless, our findings that the number of self-deducted points was affected by guilt and relational utility implied that the participants did understand how to make use of this operation. Secondly, the present studies focus only on financial self-punishment. As Tanaka et al. (2015) suggested, in addition to inflicting a financial loss on themselves, the wrongdoers could also punish themselves by suffering physical pain. Although the cost-benefit perspective implies that physical self-punishment would be affected by benefits as well, one may argue differently according to the taboo trade-off hypothesis, which suggests that body and money are incommensurable, and the sacred value of the body would be desecrated by being weighed against money (Fiske & Tetlock, 1997). Even if some wrongdoers insist on making a profit through physical self-punishment, the strong negative emotion triggered by the taboo trade-off itself could cause uncertainty in wrongdoers' behavior (Fiske & Tetlock, 1997). Therefore, it is difficult to predict whether or how relational utility influences physical self-punishment. It is an interesting topic for future studies.

In summary, self-punishment is not as empathic and moral as it appears. The present studies demonstrate that relational utility affects self-punishment. In both direct and indirect reciprocity contexts, people strategically use self-punishment to pursue profits through the repair of damaged relationships. These results help us understand self-punishment in real life.

Acknowledgments

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 Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (15XNLQ05).

References

Bastian, B., Jetten, J., & Fasoli, F. (2011). Cleansing the soul by hurting the flesh: The guilt-reducing effect of pain. *Psychological Science, 22*, 334–335. doi: 10.1177/0956797610397058

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*, 59–78. doi: 10.1017/S0140525X11002202

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. doi: 10.3758/BF03193146

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

Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress the spheres of justice. *Political Psychology, 18*, 255–297. doi: 10.1111/0162-895X.00058

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. Hill Goldsmith (Eds), *Handbook of affective sciences* (pp. 852–870). Oxford, UK: Oxford University Press.

Inbar, Y., Pizarro, D. A., Gilovich, T., & Ariely, D. (2013). Moral masochism: On the connection between guilt and self-punishment. *Emotion, 13*, 14–18. doi: 10.1037/a0029749

Nelissen, R. M. A. (2011). Guilt-induced self-punishment as a sign of remorse. *Social Psychological and Personality Science, 3*, 139–144.

Nelissen, R. M. A. (2014). Relational utility as a moderator of guilt in social interactions. *Journal of Personality and Social Psychology, 106*, 257–271. doi: 10.1037/a0034711

Nelissen, R. M. A., & Zeelenberg, M. (2009). When guilt evokes self-punishment: Evidence for the existence of a Dobby Effect. *Emotion, 9*, 118–122. doi: 10.1037/a0014540

Nowak, M. A., & Sigmund, K. (1998). The dynamics of indirect reciprocity. *Journal of Theoretical Biology, 194*, 561–574. doi: 10.1006/jtbi.1998.0775

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*, 114–123. doi: 10.1016/j.evolhumbehav.2008.09.004

Ohtsubo, Y., & Yagi, A. (2015). Relationship value promotes costly apology-making: Testing the valuable relationships hypothesis from the perpetrator's perspective. *Evolution and Human Behavior, 36*, 232–239. doi: 10.1016/j.evolhumbehav.2014.11.008

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*, 1–7. doi: 10.1037/ebs0000016

Tangney, J. P., & Dearing, R. L. (2003). *Shame and guilt* (pp. 78–89). New York, NY: Guilford Press.

Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*, 35–57. doi: 10.1086/406755

Watanabe, E., & Ohtsubo, Y. (2012). Costly apology and self-punishment after an unintentional transgression. *Journal of Evolutionary Psychology, 10*, 87–105. doi: 10.1556/JEP.10.2012.3.1

Received March 31, 2016
Revision received September 6, 2016
Accepted September 7, 2016
Published online March 15, 2017

Chao Liu
State Key Laboratory of Cognitive Neuroscience and Learning
Beijing Normal University
No. 19 Xinjiekouwai Street
Beijing, 100875
PR China
liuchao@bnu.edu.cn

Xiaoqin Mai
Department of Psychology
Renmin University of China
No. 59 Zhongguancun Street
Beijing, 100872
PR China
chinamaixq@ruc.edu.cn