frontiers
in Psychology

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fpsyg.2021.676064

Third-Party Punishment or Compensation? It Depends on the Reputational Benefits

Zhuang Li[1], *Gengdan Hu*[1,2]*, *Lei Xu*[3] *and Qiangqiang Li*[1]

[1] *Department of Psychology, School of Humanities, Tongji University, Shanghai, China,* [2] *School of Medicine, Shanghai Pudong New Area Mental Health Center, Shanghai, China,* [3] *Electrical College, Shanghai Dianji University, Shanghai, China*

Third-party fairness maintenance could win some reputational benefits, and it includes two methods: punishment and compensation. We predicted that the third parties' preference between punishment and compensation are affected by whether they are free to choose between the two methods, and the affection could be interpreted through reputational benefits. The present study includes two sections. In Study 1, the participants acted as fourth parties who were asked to rate the reputations of the third parties who had chosen different response methods to an unfair result of the dictator game. The results showed that (1) there was no reputational difference between the two methods when third parties were not free to choose, (2) but the reputation of compensation was better when third parties were free to choose. In Study 2, the participants acted as third parties. The participants were asked to choose a method to respond to an unfair result of the dictator game. There were two reputational contexts: secret and open. The results showed that (1) when third parties were not free to choose, they had no preference between the two methods under the two reputational contexts, (2) but when third parties were free to choose freely, they prefer punishment under the secret context but prefer compensation under the open context. This study systematically reveals a reputational interaction between fourth and third parties, and verifies the affection of reputational benefits on the third parties' preference between punishment and compensation.

Keywords: third-party fairness maintenance, fourth party, reputation, punishment, compensation

OPEN ACCESS

Edited by:
Matjaž Perc,
University of Maribor, Slovenia

Reviewed by:
Valerio Capraro,
Middlesex University, United Kingdom
Xiaojie Chen,
University of Electronic Science and Technology of China, China

****Correspondence:***
Gengdan Hu
hugengdan@tongji.edu.cn

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *04 March 2021*
Accepted: *07 April 2021*
Published: *28 May 2021*

Citation:
Li Z, Hu G, Xu L and Li Q (2021) Third-Party Punishment or Compensation? It Depends on the Reputational Benefits. Front. Psychol. 12:676064. doi: 10.3389/fpsyg.2021.676064

INTRODUCTION

Fairness is an important social rule. Many studies have found that to maintain fairness, people often intervene in unfair events to which they are not connected, and this is referred to as third-party fairness maintenance (Fehr and Gächter, 2002; Fehr and Fischbacher, 2003, 2004; Kroupa, 2014; Hu et al., 2015; Liu Y. et al., 2019). Compared with second-party fairness maintenance (Yamagishi et al., 2009), third-party fairness maintenance would be more neutral and impartial (Bendor and Swistak, 2001). Thus, in our everyday lives, it is often the third parties (such as the police and the courts) who typically take the responsibility of maintaining fairness.

The Reputational Benefit of Third-Party Fairness Maintenance

Third-party fairness maintenance is beneficial for a group's interests (Fehr and Gächter, 2002). However, there is a dilemma that cannot be ignored: it will lead to an individual cost. If such dilemma continues, third parties involved in fairness maintenance would eventually be eliminated in the process of evolution, but in fact, third-party fairness maintenance is widespread in both the laboratory and the field (Fehr and Gächter, 2002; Fehr and Fischbacher, 2003, 2004; Kroupa, 2014; Hu et al., 2015; Liu Y. et al., 2019). What, therefore, is the motivation for and evolutionary mechanism of third-party fairness maintenance? The indirect reciprocity theory holds that people pay some price to maintain fairness because doing so would win them a good reputation, which will bring some benefit to them in future interactions, and the reputational benefits may cover the costs. This thereby improves the adaptability of third parties (Alexander, 1986; Fehr and Fischbacher, 2003; Bereczkei et al., 2007). Empirical studies have also found that, when in the face of unfair events, compared to bystanders, third parties involved in fairness maintenance are evaluated as more trustworthy, have more opportunities to be chosen as partners, receive more material rewards than bystanders, and are more likely to be chosen as leaders (Barclay, 2006; Santos et al., 2013; Gordon et al., 2014; Raihani and Bshary, 2015a; Jordan et al., 2016).

The premise that third-party fairness maintenance can bring indirect reciprocity is due to the reputation. Whether third-party fairness maintenance results in one having a good reputation largely depends on oberservers. Only when third-party fairness maintenance is seen by fourth-party observers can it win a good reputation. It has been found that when there are audience present, and even only when experimenters were present, third-party fairness maintenance increases (Kurzban et al., 2007). On the contrary, under the condition of secrecy or anonymity, third-party fairness maintenance decreases (Burnham, 2003; Piazza and Bering, 2008). These findings indicate that the reputational context is an important factor that can affect whether a third party would intervene in unfair events.

Third-Party Punishment and Compensation

In the face of unfair events, third-party fairness maintenance has two intervention methods: punishment and compensation. Punishment is aimed at violators and related to anger (Fehr and Gächter, 2002; Lotz et al., 2011; Liu et al., 2017; Rodrigues et al., 2018), while compensation is aimed at victims and related to compassion (Condon and Desteno, 2011; Leliveld et al., 2012; Sierksma et al., 2014; Hu et al., 2015; Rodrigues et al., 2018). These methods both indicate a third party's concern for fairness and group interests, and both would lead to some good reputation, but there are still some reputational differences between the two methods.

The reputation of punishment may be more complicated than that of compensation. Barclay (2006) divided reputation into four subdimensions: niceness, trustworthiness, group concernedness, and respectability. He found that compared with non-interveners, punishers tended to be better evaluated in the latter three dimensions but were considered less nice. In addition, not all punishments win a good reputation; only those punishments that had been applied against violators would lead to a good reputation. Punishments against cooperators would lead to a bad reputation; however, all forms of compensation, even those for non-victims, would lead to a good reputation (Ozono and Watabe, 2012). Gordon et al. (2014) found that only a dominant punisher can obtain a good reputation because a dominant punisher can prevent revenge.

In addition, many studies have found that, compared with punishment, compensation may be more easily rewarded. For example, Raihani and Bshary (2015a) found that compared with non-intervention, third-party punishment can lead to monetary rewards, but third-party compensation can be rewared more. In economic activities, when interacting with punishers, people are more willing to expect to receive resources from punishers rather than providing resources to them (Horita, 2010). Ozono and Watabe (2012) had made similar observations. They provided their participants with three candidates as partners in the economic game: the punisher, the compensator, and the non-intervener. It was found that people preferred to choose the compensator and non-intervener as partners rather than the punisher. In the dictator game, when the participants were dictators, they allocated less money to the punisher but allocated more to the compensator.

Why are there differences between the two methods regarding reputational benefits? Punishment and compensation not only convey the third parties' concern about fairness but also convey the information about their characteristic traits. Punishment can give people an impression of emotional instability and irritability, which can lead to others becoming feared; however, compensators can make people feel gentle and empathetic (Gordon et al., 2014; Kroupa, 2014; Raihani and Bshary, 2015b). Clearly, the latter would be more preferable. In addition, people need to infer the motivation for third-party fairness maintenance according to the intervention method that had been applied. Only when motivation is really out of concern for fairness and group interests can people provide good reputational feedback for third parties (Raihani and Bshary, 2015b). Punishment may be due to the emotional venting of third parties, which would greatly reduce their altruistic attributes, and then negatively impact their reputations.

How should people infer the characteristic traits and motivation behind the two intervention methods of punishment and compensation? Whether third parties are free to choose between these methods may provide a clue. When exploring punishment and compensation, researches usually involve two types of settings: one is that a third party chooses between punishment, compensation, and keep (i.e., to not intervene); the other is that they choose between punishment and keep or between compensation and keep (Chavez and Bicchieri, 2013; Rodrigues et al., 2018; Liu Y. et al., 2019). In the former case, third parties are free to choose between the two intervention methods, while in the latter case, they are not. When third parties are not free to choose, both the methods indicate a third party's willingness to maintain fairness and a third party's concern for fairness and group interests; however, it does not reflect the third party's preference between the two intervention methods. However, when third parties can freely choose, their choices can reflect their preference between the two methods and further reflect their characteristic traits and true motivation. Therefore, we put forward the following as our first hypothesis:

H1: When third parties are unable to freely choose between punishment and compensation, there is no difference in the reputation obtained between the two methods; however, when third parties are able to freely choose, the reputation of compensation is better than that of punishment.

Whether or not one can freely choose between punishment or compensation cannot only affect the reputation of the two methods but also the third parties' choice between the two methods. Chavez and Bicchieri (2013) found that when third parties can only punish, third parties tend to punish, but when third parties can choose between punishment and compensation, third parties tend to compensate. It is thus evident that third parties intentionally conduct reputation management by choosing between these two intervention methods. When punishment is the only way, third parties may show their concern for fairness and group interests through a preference for punishment. When they are free to choose, third parties show their concern for fairness and group interests as well as the empathy traits through a preference for compensation.

The reputation management is also reflected in the fact that third parties will attempt to avoid the damage to reputation caused by punishment. For example, Rockenbach and Milinski (2011) found that third-party punishers even spend money to hide their punitive behavior. Horita and Takezawa (2014) argued that the presence of audience may be a reference clue for third parties to conduct punishment because they found that the presence of audience only enhances third-party punishment among the people who were not hot tempered but not the people who were hot tempered. It was because third parties who are hot tempered generally do not want their anger to be discovered by others, and due to this, they tend to hide their anger by reducing punishments. Based on the above review, we propose our second hypothesis:

H2: Third parties' preference between punishment and compensation is affected by reputational context. Specifically, under the context of open, which means that the method of third-party fairness maintenance is known by others, if third parties can choose between punishment and compensation, they tend to compensate; if third parties are unable to freely choose, there is no preference between the two methods. Under the context of "secret," which means that the method of third-party fairness maintenance is not known by others, whether third parties can choose between the two methods or not, there is no preference.

The Present Study

This study has two focuses: one is the reputation evaluation of fourth parties on third-party fairness maintenance, and the other is the response of third parties to unfair events under the two different reputation contexts of open and secret; these two aspects are both related to whether third parties have the right to free choice. This study includes two component studies. In Study 1, our participants acted as fourth parties and were asked to rate the reputation of third-party fairness maintenance. Two experiments were conducted. In Study 1a, the third parties were not free to choose between punishment and compensation, while in Study 1b, the third parties were free to choose. In Study 2, our participants acted as third parties and were asked to respond to unfair events. Study 2 also included two experiments: in Study 2a, the third parties could not freely choose between the two intervention methods, while in Study 2b, they could freely choose.

STUDY 1: FOURTH PARTIES' REPUTATION EVALUATION OF THIRD-PARTY FAIRNESS MAINTENANCE

Study 1a: Third Parties Were Not Free to Choose

In this experiment, the participants acted as fourth parties. We presented to them a scenario that involved the dictator game and third-party fairness maintenance. In the scenario, the third parties had to choose between the methods of punishment and compensation as a response to an unfair allocation scheme of the dictator game, and the third parties could not freely choose between the two methods. The participants were asked to evaluate the reputation of the third parties according to the latter's chosen response method. This experiment was concerned about whether there is a difference in the reputation derived from the two methods when a third party is not free to choose between them.

Participants

A total of 103 college students were recruited from Shanghai Dianji University (male, 44; age range, 17–23; mean age, $M \pm SD = 19.42 \pm 1.40$). The study was approved by the academic ethics committee of Shanghai Dianji University. All the participants voluntarily participated and signed written informed consent forms. The participants were promised that the experimental results would be kept confidential and would only be used for academic research.

Procedure

We presented a scenario that involved the dictator game and third-party fairness maintenance to the fourth-party participants. To increase the authenticity of the experiment, we informed the participants that the scenario had actually happened in our previous experiment. The scenario was as follows: 10 RMB (∼US$ 1.55) was allocated between two people, and one of them (dictator) proposed an allocation plan, while the other (recipient) had to accept it. As a result, the dictator allocated 8 RMB (∼US$ 1.24) to himself and only 2 RMB (∼US$ 0.31) to the recipient.

The third parties needed to respond to this unfair event. The third parties were divided into two treatments; in one, the third parties had to choose between punishment and keep (not intervene), while in the other, the third parties had to choose between compensate and keep.

In both treatments, the third parties were endowed with 10 RMB. If a third party chose to keep, all the 10 RMB would be left to the third party. If the third party chose to punish/compensate, they would have to further decide how much money would be transferred as punishment/compensation. For each RMB to punish/compensate, the dictator/recipient lost/received 1 RMB (∼US$ 0.15). The transfer amount of punishment/compensation was sourced from the initial 10 RMB; hence, the choice of punishment/compensation caused some losses for the third party (It should be noted that in the relevant studies, the transfer amount and the lose/gain amount were not equal. Usually, for each RMB to punish/compensate, the dictator/recipient loses/receives 3 RMB. In the present experiment, we were only concerned about what the third parties spend, not what the violators'/victims' lose/gain. In order not to increase the difficulty of arithmetic comprehension, we set the amount spent and the amount lost/gained to be equal). In each treatment, we presented the participants with three choices regarding the third parties, which led to six choice schemes:

(1) Choose keep in the punishment treatment (no punishment).
(2) Choose punishment in the punishment treatment, with a transfer amount of 3 RMB (∼US$ 0.46).
(3) Choose punishment in the punishment task, with a transfer amount of 5 RMB (∼US$ 0.77).
(4) Choose keep in the compensation treatment (no compensation).
(5) Choose compensation in the compensation treatment, with a transfer amount of 3 RMB.
(6) Choose compensation in the compensation treatment, with a transfer amount of 5 RMB.

We asked the participants to rate the reputation of the third parties of the above six cases from four dimensions: niceness, trustworthiness, group concernedness, and respectability (Barclay, 2006). The Cronbach α was 0.744 in this study. The score range of each dimension was between 1 and 7, and the range of the total score was between 4 and 28. The higher the score, the better the reputation. The experiment was conducted using a computer with the software Eprime 2.0. There were 24 trials; they were presented randomly.

Results

If the participants chose "keep," the transfer amount of punishment/compensation was regarded as 0. We used a 2 (method: punishment, compensation) × 3 (amount: 0, 3, 5) repeated measures analysis of variance (ANOVA) to compare the participants' total scores (see **Table 1** and **Figure 1**). The results showed that the main effect of the method was not significant, $F_{(1,\,102)} = 0.58$, $p = 0.45$, $\eta^2{}_p = 0.006$, which indicates that there was no difference in the reputation between the two methods. The main effect of the amount was significant, $F_{(2,\,204)} = 988.57$, $p < 0.001$, $\eta^2{}_p = 0.91$. The higher the amount, the higher the total score, and a *post-hoc* test showed that the difference between two amounts was significant. The interaction effect was significant, $F_{(2,\,204)} = 221.56$, $p < 0.001$, $\eta^2{}_p = 0.69$. The simple effect test showed that the total score of punishment was significantly higher than that of compensation when the amount was 0, but compensation was significantly higher than punishment when the amount was 5.

Previous studies have found that there are differences in specific dimensions of reputation between the two methods (Barclay, 2006). We performed a 2 (method: punishment, compensation) × 3 (amount: 0, 3, 5) repeated measures ANOVA four times to compare the reputational scores for each specific dimension (see **Table 1** and **Figure 1**).

In the niceness dimension, the main effects of the method and amount were both not significant: method, $F_{(1,\,102)} = 0.74$, $p = 0.39$, $\eta^2{}_p = 0.01$; amount, $F_{(2,\,204)} = 3.06$, $p = 0.05$, $\eta^2{}_p = 0.03$. However, the interaction effect between the two was significant, $F_{(2,\,204)} = 571.88$, $p < 0.001$, $\eta^2{}_p = 0.85$. The simple effect test showed that, in punishment, the higher the amount was, the lower the score for niceness was, while in compensation, the higher the amount was, the higher the score for niceness was. This indicates that third-party fairness maintenance was considered to be less nice in the punishment treatment but nicer in the compensation treatment.

For the other three specific dimensions, the main effects of the method were all not significant: trustworthiness, $F_{(1,\,102)} = 0.48$, $p = 0.49$, $\eta^2{}_p = 0.05$; group concernedness, $F_{(1,\,102)} = 1.86$, $p = 0.18$, $\eta^2{}_p = 0.02$; respectability, $F_{(1,\,102)} = 1.60$, $p = 0.21$, $\eta^2{}_p = 0.02$. This indicates that there was no reputational difference in the three specific dimensions between the two methods. The main effects of the amount in the three specific aspects were all significant: trustworthiness, $F_{(2,\,204)} = 572.27$, $p < 0.001$, $\eta^2{}_p = 0.85$; group concernedness, $F_{(2,\,204)} = 741.48$, $p < 0.001$, $\eta^2{}_p = 0.88$; respectability, $F_{(2,\,204)} = 710.02$, $p < 0.001$, $\eta^2{}_p = 0.85$. The higher the amount, the higher the reputational scores in the three dimensions. The interaction effects were all not significant: trustworthiness, $F_{(2,\,204)} = 1.14$, $p = 0.32$, $\eta^2{}_p = 0.01$; group concernedness: $F_{(2,\,204)} = 2.06$, $p = 0.13$, $\eta^2{}_p = 0.02$; respectability, $F_{(2,\,204)} = 3.14$, $p = 3.14$, $\eta^2{}_p = 0.03$. This indicates that whether punishment or compensation, the reputation of third-party fairness maintenance is better than that of keep in these three dimensions. In addition, the higher the transfer amount, the better the reputation one may have.

From the above results, we can see that when third parties cannot choose freely, there will be no difference in the reputation between punishment and compensation, which is consistent with our hypothesis. Whether punishment or compensation, the higher the transfer amount, the better the reputation, but in the niceness dimension, the opposite was true, which indicates that punishment is considered to be a less preferable method to maintain fairness.

Study 1b: Third Parties Were Free to Choose

Similar to Study 1a, in Study 1b, the participants still acted as the fourth parties. Unlike the first study, the participants were free to choose between punishment and compensation. The experiment was concerned about whether there is a difference in the reputation between the two methods when third parties are free to choose between them.

Participants

A total of 107 college students were recruited from Shanghai Dianji University (male, 54; age range, 17–24; mean age,

TABLE 1 | The reputational score when third parties were not free to choose (*M* ± *SD*).

Score	Punishment			Compensation		
	0 (keep)	**3**	**5**	**0 (keep)**	**3**	**5**
Total	11.00 ± 2.10	14.71 ± 2.97	18.57 ± 1.99	7.82 ± 1.80	15.07 ± 2.41	21.72 ± 2.08
Niceness	5.03 ± 1.01	3.72 ± 0.92	2.09 ± 0.97	2.01 ± 0.66	3.67 ± 0.95	5.34 ± 0.88
Trustworthiness	1.97 ± 0.75	3.75 ± 0.86	5.38 ± 0.88	2.08 ± 0.85	3.82 ± 0.93	5.30 ± 0.82
Group-concernedness	1.97 ± 0.86	3.67 ± 0.95	5.47 ± 0.81	1.89 ± 0.69	3.89 ± 1.05	5.56 ± 0.86
Respectability	2.03 ± 0.90	3.57 ± 1.12	5.64 ± 0.77	1.83 ± 0.72	3.69 ± 0.93	5.51 ± 0.86

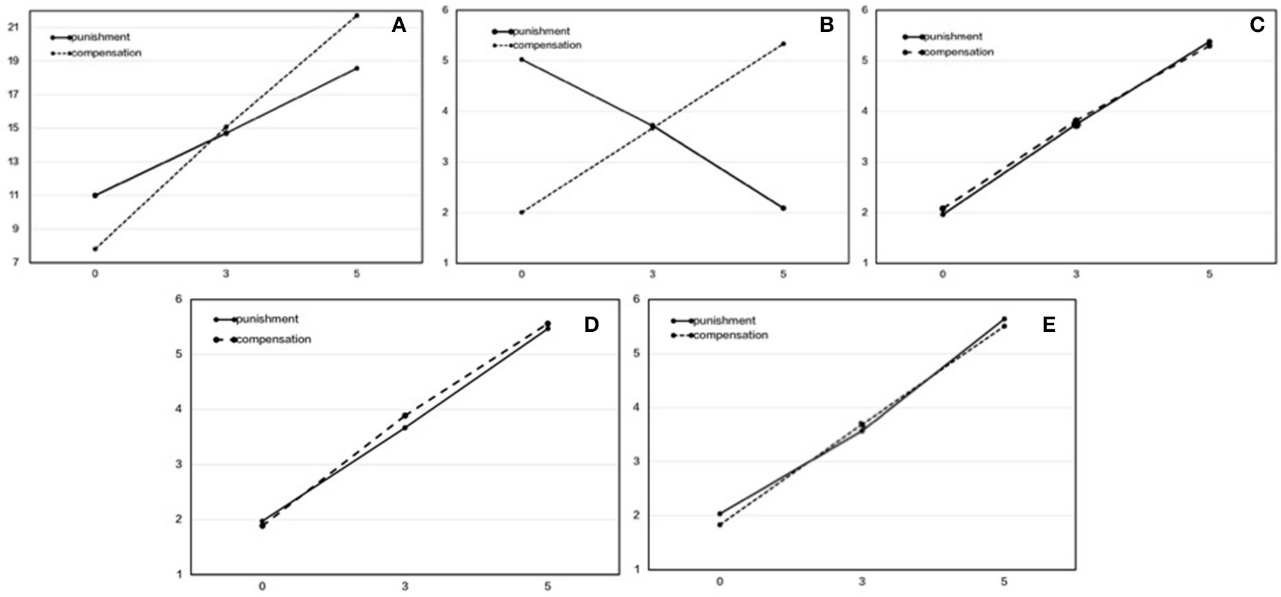

FIGURE 1 | The difference in reputation between punishment and compensation when third parties were not free to choose. The horizontal axis is the transfer amount, and the longitudinal axis is the reputational score. **(A)** Total score. **(B)** Niceness. **(C)** Trustworthiness. **(D)** Group-concernedness. **(E)** Respectability.

M ± *SD* = 19.19 ± 1.41). The study was approved by the academic ethics committee of Shanghai Dianji University. All the participants voluntarily participated and signed written informed consent forms. The participants were promised that the experimental results would be kept confidential and that they would only be used for academic research.

Procedure

The participants were presented with a scenario which was similar to Study 1a, and they had to choose a respond method. Unlike Study 1a, there was only one case for the third parties; they could choose between keep (not intervene), punishment, and compensation, which means that the third parties were free to choose between compensation and punishment. We presented the participants with five choices regarding third parties:

(1) Choose keep (non-punishment and non-compensation).
(2) Choose punishment, with a transfer amount of 3 RMB.
(3) Choose punishment, with a transfer amount of 5 RMB.
(4) Choose compensation, with a transfer amount of 3 RMB.
(5) Choose compensation, with a transfer amount of 5 RMB.

Similar to Study 1a, the participants had to rate the reputation in the five subdimensions (Barclay, 2006). The Cronbach α of the five subdimensions in the study was 0.767. There were 20 trials in this study, which were presented randomly.

Results

We treated the choice of "keep" as non-intervention and treated the punishment of 3 RMB and 5 RMB as punishment and the compensation of the two amounts as compensation. Thus, the third parties were divided into three treatments (keep, punishment, and compensation). We used a repeated measures ANOVA to compare the total reputational scores of the three treatments. The total reputational score of punishment/compensation was the average score of the two amounts. The results showed that the difference was significant, and a *post-hoc* test showed that the compensation was higher than punishment and keep, and punishment was significantly higher than compensation. Following this, we conducted a repeated measures ANOVA to compare the reputational score of the four specific dimensions in the three treatments. The results showed that the differences in the four specific dimensions were

TABLE 2 | The difference in the reputational scores when third parties were free to choose (*M* ± *SD*).

Score	Keep	Punishment	Compensation	$F_{(2, 212)}$	η^2_p
Total	$11.58^c \pm 3.40$	$19.02^b \pm 2.02$	$24.34^a \pm 1.82$	1021.84***	0.91
Niceness	$5.17^b \pm 0.99$	$3.70^c \pm 0.81$	$6.01^a \pm 0.54$	270.54***	0.72
Trustworthiness	$2.12^c \pm 1.09$	$5.18^b \pm 0.72$	$6.05^a \pm 0.64$	828.04***	0.89
Group-concernedness	$2.25^c \pm 1.21$	$4.83^b \pm 0.64$	$6.13^a \pm 0.56$	649.35***	0.86
Respectability	$2.04^c \pm 1.19$	$5.32^b \pm 0.72$	$6.14^a \pm 0.60$	773.22***	0.88

****p < 0.001. a was higher significantly than b and c, and b was higher significantly than c.*

TABLE 3 | The reputational scores of different transfer amounts for punishment and compensation (*M* ± *SD*).

Score	Punishment		Compensation	
	3	5	3	5
Total	19.33 ± 2.49	18.73 ± 2.93	22.58 ± 3.05	26.09 ± 1.48
Niceness	4.46 ± 1.07	2.93 ± 1.11	5.47 ± 0.93	6.56 ± 0.54
Trustworthiness	5.17 ± 0.96	5.19 ± 1.03	5.64 ± 0.94	6.47 ± 0.68
Group-concernedness	4.47 ± 0.97	5.21 ± 1.00	5.70 ± 0.86	6.55 ± 0.59
Respectability	5.23 ± 1.09	5.41 ± 0.95	5.78 ± 1.14	6.51 ± 0.57

all significant. In the niceness dimension, compensation was higher than punishment and keep, and keep was higher than punishment, while in the other three dimensions, compensation was higher than punishment and keep, and punishment was significantly higher than keep (see **Table 2**).

To further investigate the reputational difference of punishment and compensation between the different amounts, we used a 2 (amount: 3, 5) × 2 (method: punishment, compensation) repeated measures ANOVA five times to compare the differences in the total scores and the four specific dimensions (see **Table 3** and **Figure 2**).

Regarding the total scores, the main effect of the method was significant, $F_{(1,\,106)} = 503.07$, $p < 0.001$, $\eta^2{}_p = 0.83$, and compensation was higher than punishment. This indicates that the reputation of compensation is better than that of punishment. The main effect of the amount was significant, $F_{(1,\,106)} = 53.69$, $p < 0.001$, $\eta^2{}_p = 0.34$, and the reputational score for 5 RMB was higher than that for 3 RMB. This indicates that the higher the transfer amount, the better the reputation. The interaction effect was significant, $F_{(1,\,106)} = 62.75$, $p < 0.001$, $\eta^2{}_p = 0.37$. A simple effect test showed that, in punishment, the reputational score for 3 RMB was higher than that for 5 RMB, while in compensation, the reputational score for 5 RMB was higher than that for 3 RMB.

In the niceness dimension, the main effect of the method was significant, $F_{(1,\,106)} = 590.34$, $p < 0.001$, $\eta^2{}_p = 0.85$. Compensation was higher than punishment. The main effect of the amount was significant, $F_{(1,\,106)} = 7.43$, $p = 0.007$, $\eta^2{}_p = 0.07$. In addition, 3 RMB was higher than 5 RMB. The interaction effect was significant, $F_{(1,\,106)} = 207.46$, $p < 0.001$, $\eta^2{}_p = 0.66$. A simple effect analysis showed that, in punishment, the reputational score for 3 RMB was higher than that for 5 RMB, while in compensation, that for 5 RMB was higher than that for 3 RMB.

In the other three dimensions, the main effects of the method were all significant, and the scores for compensation were higher than those for punishment: trustworthiness, $F_{(1,\,106)} = 104.29$, $p < 0.001$, $\eta^2{}_p = 0.50$; group concernedness, $F_{(1,\,106)} = 232.86$, $p < 0.001$, $\eta^2{}_p = 0.69$; respectability, $F_{(1,\,106)} = 96.95$, $p < 0.001$, $\eta^2{}_p = 0.48$. The main effects of the amount were all significant, and the scores for 5 RMB were higher than those for 3 RMB: trustworthiness, $F_{(1,\,106)} = 26.84$, $p < 0.001$, $\eta^2{}_p = 0.20$; group concernedness, $F_{(1,\,106)} = 91.21$, $p < 0.001$, $\eta^2{}_p = 0.46$; respectability, $F_{(1,\,106)} = 28.30$, $p < 0.001$, $\eta^2{}_p = 0.21$. The interaction effect of trustworthiness was significant, $F_{(1,\,106)} = 24.01$, $p < 0.001$, $\eta^2{}_p = 0.19$. The simple effect test showed that, in punishment, the difference was not significant between 3 RMB and 5 RMB, while in compensation, that for 5 RMB was significantly higher than that for 3 RMB. The interaction of group concernedness was not significant, $F_{(1,\,106)} = 0.41$, $p = 0.53$, $\eta^2{}_p = 0.004$. The interaction effect of respectability was significant, $F_{(1,\,106)} = 7.30$, $p = 0.008$, $\eta^2{}_p = 0.064$. The simple effect test showed that, in punishment, the difference was not significant between 3 RMB and 5 RMB, while in compensation, that for 5 RMB was significantly higher than that for 3 RMB.

From the above results, we can see that no matter the total score or subdimensions, the reputation of compensation is better than that of punishment, which indicates that compensation can lead to a better reputation when third parties are free to choose between the two methods. This is different from Study 1a but consistent with our hypothesis. Moreover, compared with high punishment, low punishment is considered to be nicer, which indicates that punishment is considered to be less nice. This is consistent with the findings of Study 1a.

STUDY 2: THIRD PARTIES' RESPONSE TO UNFAIR EVENTS

Study 2a: Third Parties Were Not Free to Choose

In this study, the participants acted as the third parties. We presented them with unfair allocation schemes of the dictator game, and there were two reputational contexts: open and secret. In the face of unfair events, the third parties either chose to respond between keep and punish or between keep and compensate. We were concerned about whether reputational contexts would affect the choice between the two methods when third parties are not free to choose.

Participants

A total of 225 college students were recruited from Shanghai Dianji University (male, 126; age range, 16–24; mean age, *M* ± *SD* = 19.86 ± 1.75). The study was approved by the academic ethics committee of Shanghai Dianji University. All the participants voluntarily participated and signed written

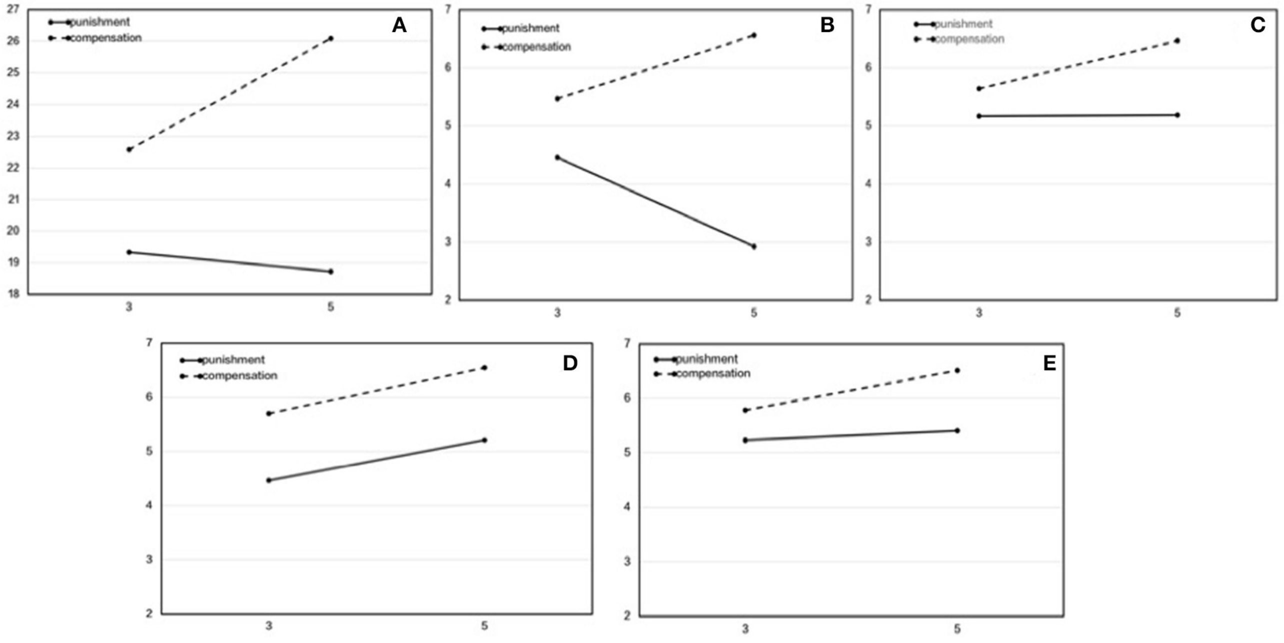

FIGURE 2 | The difference in reputation between punishment and compensation when third parties were free to choose. The horizontal axis is the transfer amount, and the longitudinal axis is the reputational score. **(A)** Total score. **(B)** Niceness. **(C)** Trustworthiness. **(D)** Group-concernedness. **(E)** Respectability.

informed consent forms. The participants were promised that the experimental results would be kept confidential and that they would only be used for academic research.

Procedure

The experiment was carried out using a computer with the software Eprime 2.0. The participants were taken to a separate room where they were told that there were two other people playing the dictator game in two other rooms: one person was the dictator, and the other was the recipient. In the game, the dictator had to allocate 10 RMB between himself and the recipient. There were nine allocation schemes for the dictator to choose: 9:1 (i.e., the dictator was allocated 9 RMB and the recipient 1 RMB), 8:2, 7:3, 6:4, 5:5, 4:6, 3:7, 2:8, and 1:9. The dictator's choice was shown to the participants via a computer, and we presented five allocation schemes to the participants: 9:1, 8:2, 7:3, 6:4, and 5:5. However, the dictator game did not, in fact, occur. The five presented schemes had been preset via the computer programs, but the participants thought that they had been made by a dictator. The participants were randomly divided into four experimental treatments of 2 (reputational context: open, secret) × 2 (method: punishment, compensation): open punishment, open compensation, secret punishment, and secret compensation.

Under the open context, the participants were informed that their choices would be open to other participants when the experiment ended; however, this did not occur, but the participants believed it would. Under the secret context, the participants were informed that their choices would be kept secret. In the punishment condition, the participants chose between keep and punishment, while in the compensation condition, they chose between keep and compensation.

The participants were endowed with a basic fee of RMB 10 (∼US$1.55) and another additional 50 money units (MUs); each MU corresponded to RMB 0.1 (∼US$0.016). If the participants chose keep, all the basic fees and additional MUs, a total of RMB 15 (∼US$2.33), would be left to themselves. If they chose to punish/compensate, they would further select how much MUs would be transferred to punish/compensate. There were four transfer schemes that participants could select: 5, 10, 15, and 20. For each MU that would be transferred as a punishment/compensation, the dictator/recipient would lose/gain 3 MUs. The transferred MUs were obtained from the original 50 MUs. There were five trials that were presented randomly. We choose one trial randomly and paid the test fees according to the participants' selections. The test fee was calculated as follows: 10 + (50-transfer MUs)/10.

Results

Previous studies have found that in the face of fair allocation, third parties generally chose keep (Hu et al., 2015; Liu Y. et al., 2019). In this experiment, for the 5:5 allocation scheme, all the participants chose keep. Based on previous research, we only analyzed four unfair schemes (Hu et al., 2015; Liu Y. et al., 2019).

The choice of keep means non-intervention, and the choice of punish and compensate should be treated as an intervention. Thus, the expected rate for keep and intervention were all 0.5. To investigate whether third parties would be inclined to intervene in unfair events under four experimental conditions, we used a single sample *t*-test to compare the intervention rates with the rate of 0.5 (see **Table 4**). Intervention rates were calculated as the intervention times divided by 4 (Liu Y. et al., 2019). The results showed that in the four conditions, the intervention rates were all significantly higher than 0.5: secret punishment, $t_{(224)} = 3.34$, $p = 0.002$; secret compensation, $t_{(224)} = 2.56$, $p = 0.01$; open punishment, $t_{(224)} = 6.65$, $p < 0.001$; and open compensation, $t_{(224)} = 8.14$, $p < 0.001$. This indicates that, when faced with unfair events, third parties are inclined to intervene.

To investigate the effect of the reputational contexts and the intervention method on transfer rates, we used a two-way ANOVA of 2 (reputational context: secret, open) × 2 (method: punishment, compensation) to compare the transfer rates in the four experimental treatments (see **Figure 3**). The results showed that the main effect of the reputational context was significant, $F_{(1,\,221)} = 8.04$, $p = 0.005$, $\eta^2{}_p = 0.04$, and that of open was significantly higher than that of secret. The main effect of the method was not significant, $F_{(1,\,221)} = 0.50$, $p = 0.48$, $\eta^2{}_p = 0.002$. Furthermore, the interaction effect was not significant, $F_{(1,\,221)} = 0.64$, $p = 0.42$, $\eta^2{}_p = 0.003$.

We then used a same two-way ANOVA to compare the transfer amounts in the four experimental treatments (see **Figure 3**). The results showed that the main effect of the reputational context was significant, $F_{(1,\,224)} = 12.11$, $p = 0.001$, $\eta^2{}_p = 0.05$, and that of open was significantly higher than that of secret. The main effect of the method was not significant, $F_{(1,\,224)} = 0.90$, $p = 0.35$, $\eta^2{}_p = 0.004$. The interaction effect was not significant, $F_{(1,\,224)} = 0.18$, $p = 0.68$, $\eta^2{}_p = 0.001$.

Based on the above results, we can see that, under the open context, which is good for one's reputation, third-party fairness maintenance will increase, while under the secret context, it may decrease. In addition, we can also conclude that when the third party cannot choose freely, there is no preference between the two methods regarding third-party fairness maintenance.

TABLE 4 | Transfer rates and amounts when third parties were not free to choose (*M* ± *SD*).

	Secret-punishment	Secret-compensation	Open-punishment	Open-compensation
Rate	0.67 ± 0.36	0.61 ± 0.30	0.75 ± 0.29	0.75 ± 0.25
Amount	8.53 ± 5.16	7.72 ± 4.23	10.35 ± 4.73	10.04 ± 3.61

Study 2b: Third Parties Were Free to Choose

In this experiment, the participants still acted as the third parties. The participants were presented with unfair results of the dictator game. Unlike Study 2a, in this experiment, when faced with unfair events, third parties could choose among keep, punish, and compensate. There were still two reputational treatments: secret and open. We were concerned about whether reputational contexts would affect third parties' choice between the two methods when they are free to choose.

Participants

A total of 126 college students were recruited from Shanghai Dianji University (male, 67; age range, 17–24; mean age, $M \pm SD = 20.11 \pm 1.61$). The study was approved by the academic ethics committee of Shanghai Dianji University. All the participants voluntarily participated and signed written informed consent forms. The participants were promised that the experimental results would be kept confidential and that they would only be used for academic research.

Procedure

Similar to Study 2a, the participants were asked to choose a method to respond to the unfair schemes of the dictator game, and the participants were divided into two experimental treatments: secret and open. Unlike Study 2a, faced with the game result, participants could either choose keep, punishment, or compensation as a response. All other settings were similar to that applied in Study 2a.

Results

Similar to Study 2a, faced with the result of 5:5, all the participants chose keep; thus, we solely analyzed the four unfair schemes (Hu et al., 2015; Liu Y. et al., 2019). We treated punishment and compensation as intervention, and the expected rate of the intervention was 0.5. To investigate whether third parties are inclined to intervene in unfair events, we used a single sample *t*-test to compare the intervention rates with the rate of 0.5. The results showed that in the two experimental treatments,

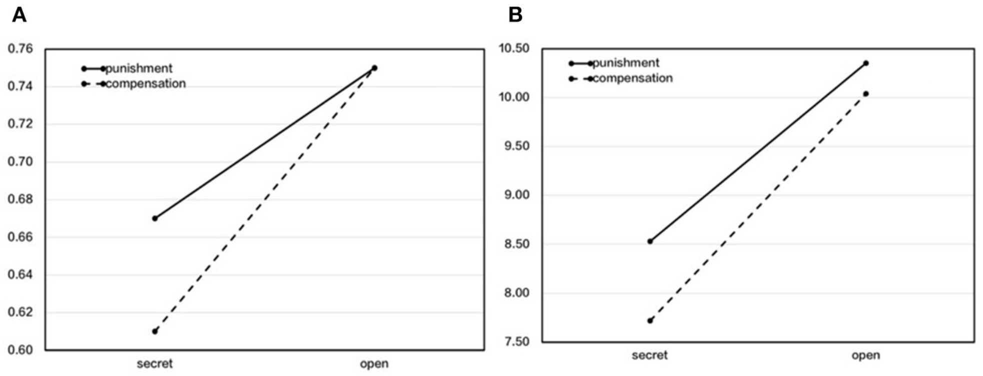

FIGURE 3 | The difference in transfer rates and amounts between punishment and compensation under the contexts of secret and open when participants were not free to choose. **(A)** Transfer rate. **(B)** Transfer amount.

TABLE 5 | Transfer rates and amounts when third parties were free to choose ($M \pm SD$).

	Secret-punishment	Secret-compensation	Open-punishment	Open-compensation
Transfer rate	0.35 ± 0.25	0.19 ± 0.23	0.23 ± 0.30	0.48 ± 0.35
Transfer amount	4.62 ± 3.88	2.36 ± 3.44	2.85 ± 3.87	6.74 ± 5.91

the intervention rates were all significantly higher than 0.5 (secret: $M \pm SD = 0.54 \pm 0.25$, $t = 1.26$, $p = 0.04$; open: $M \pm SD = 0.71 \pm 0.21$, $t = 8.01$, $p < 0.001$). This indicates whether open to others or not, third parties are inclined to intervene in unfair events.

To analyze the effect of the reputational context and the method on third-party fairness maintenance, we used a mixed two-way ANOVA of 2 (between, reputational contexts: secret, open) × 2 (within, method: punishment, compensation) to compare the intervention rates (see **Table 5** and **Figure 4**). The results showed that the main effect of the method was not significant, $F_{(1,\ 123)} = 0.69$, $p = 0.41$, $\eta^2{}_p = 0.006$. The main effect of the reputational context was significant, $F_{(1,\ 123)} = 17.60$, $p < 0.001$, $\eta^2{}_p = 0.12$, and that of open was significantly higher than that of secret. The interaction effect was significant, $F_{(1,\ 123)} = 20.18$, $p < 0.001$, $\eta^2{}_p = 0.14$. The simple effect test showed that under the secret context, the punishment rate was higher than that of compensation, while under the open context, the compensation rate was higher than that of punishment.

Following this, we used the same mixed two-way ANOVA to compare the transfer amounts (see **Table 5** and **Figure 4**). The results showed that the main effect of the method was not significant, $F_{(1,\ 123)} = 1.34$, $p = 0.24$, $\eta^2{}_p = 0.01$. The main effect of the reputational context was significant, $F_{(1,\ 123)} = 12.51$, $p = 0.001$, $\eta^2{}_p = 0.092$, and the transfer amount of open was significantly higher than that of secret. The interaction effect was significant, $F_{(1,\ 123)} = 19.86$, $p < 0.001$, $\eta^2{}_p = 0.14$. A simple effect test showed that under the secret context, the punishment amount was significantly higher than that of compensation, while under the open context, the compensation amount was significantly higher than that of punishment.

From the above results, we can see that third-party fairness maintenance increased under the open context, and this was consistent with the findings of Study 2a. What was inconsistent, which was also more important, was that we found an interaction effect. Under the secret context, the third parties tended to punish, while under the open context, they tended to compensate. This indicates that third parties' preference between punishment and compensation was mediated by the reputational context, and they conduct reputational management according to reputational contexts.

DISCUSSION

Third-party fairness maintenance will lead to a dilemma of interests between a group and an individual. Indirect reciprocity theory resolves this dilemma and provides a sound explanation regarding the motivation for the third-party fairness maintenance, and reputational benefits play an important role in this regard. The present study verified the effect of reputational benefits on third-party fairness maintenance through the dual perspectives of fourth parties and third parties.

The Interaction Focusing on the Reputation Between Fourth Parties and Third Parties

This study present us with an interesting interaction between third and fourth parties. Fourth parties evaluate the reputation through the choices of the third parties and the option, while the third party adjusts its choice according to the option and the reputational context to win a good reputation. There is a Chinese proverb: "the road is high one-foot, the evil spirit is high one unit of length," which is a vivid description of this interaction.

Reputation is a precious social resource, which can only be given to third parties who are truly altruistic. Through this, the interests of a group and third-party individuals could be guaranteed. The third parties' choice between punishment and compensation and whether they were free to choose provides a basis for the reputational evaluations of fourth parties. Third-party fairness maintenance includes two aspects: the willingness to intervene and the way to intervene. The former indicates whether a third party tends to intervene in unfair events, while the latter indicates the preference between the two methods. When third parties are free to choose, whether punishment or compensation, this means that the third parties have the willingness to maintain fairness, and there would be no reputational difference between the two methods. When third parties are free to choose, their choice can reflect the third parties' preference between the two ways and the true motivation behind it, and the reputation of compensation would be better than that of punishment. For third parties, when they are not free to choose, they need to show their concern for fairness and group interests through intervening in unfair events, while when they are free to choose, they need to show a preference for compensation and signal to others that they are concerned about both fairness and compassion.

Fourth parties typically try to understand the true motivation behind third-party fairness maintenance, while third parties try to cover up the punishment that is damaging their reputation and show the compensation that is conducive to their reputation. Fourth parties and third parties seem to be engaged in a contest. A similar phenomenon has been found in a study by Rockenbach and Milinski (2011). In their study, participants participated in the PGG with punishment, and observers were set up. They found that the observers would rather spend money to understand the players' contributions in the PGG and their punishment decisions. Players are also willing to spend to hide their punishment decisions but are also willing to open their contributions in the PGG.

In real social life, moral behaviors of different nature are often treated differently. We will protect the privacy of moral behaviors that is similar to punishment such as reporting but propagandize moral behaviors, which are similar to compensation such as donation. The whistleblowers may be retaliated, and reputational

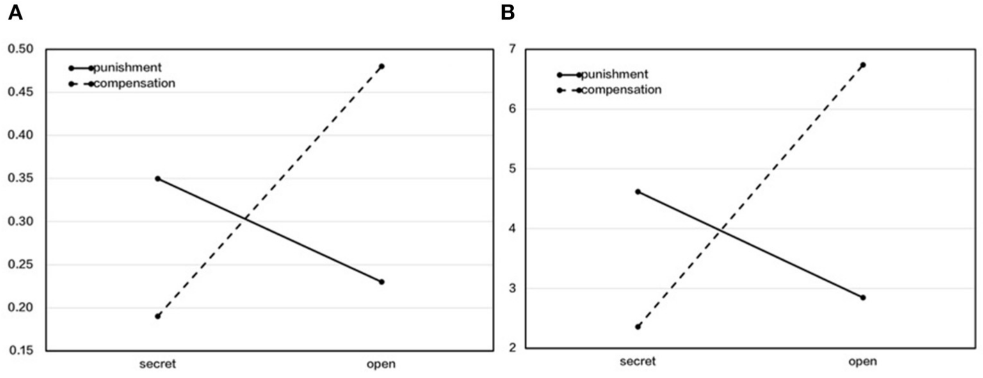

FIGURE 4 | The difference in transfer rates and amounts between punishment and compensation under the contexts of secret and open when participants were free to choose. **(A)** Transfer rate. **(B)** Transfer amount.

benefits may not be able to make up for the loss caused by retaliation. The best setting for a whistleblower is to get monetary reward and give up the reputational benefits. This is consistent with the strategy of third-party fairness maintenance: choose punishment under the secret context, but choose compensation under the open context.

Third-Party Fairness Maintenance Has a Dual Feature of Altruism and Selfishness

As a kind of prosocial behavior, third-party fairness maintenance has been labeled as moral. However, recent studies had shown that human's prosocial behaviors are not only motivated by social norms but also by personal norms (Capraro and Perc, 2021). Social norm is an external standard of moral behavior, while the personal norm is an internal standard, which will be presented in the form of moral framework (Jordan and Rand, 2020; Capraro et al., 2021). The affections of observers on the two kinds of norms are different. It has been found that people's public behavior is driven by social norms (Schram and Charness, 2015), but the private behavior is driven by personal norms (Capraro and Rand, 2018).

The present study found that in the face of unfair events, the third party will tend to intervene, but it is still affected by observers. Under the secret context, the intervention rate of third parties was higher than the theoretical rate of 0.5 and tended to punish. However, under the open context, the intervene rate of third parties was higher and tended to compensate. This indicate that third-party fairness maintenance may be driven by both social norms and personal norms. In the previous studies, both the social norms and personal norms are all altruistic. However, in this study, we found that the individual norms of third-party fairness maintenance have a consideration of self-interest.

Although it has a feature of selfish, we should not belittle the moral value of third-party fairness maintenance. Indirect reciprocity is not only the result of third-party fairness maintenance but also the motivation behind it. Third-party fairness maintenance brings indirect reciprocity, which in turn stimulates third-party fairness maintenance. In this way, a virtuous circle is formed, so that third-party fairness maintenance can be sustained, and the group interest, as well as fairness morality, can also be guaranteed.

Third-Party Punishment Is a Double-Edged Sword

In this study, we found that the punisher is considered to be less nice. Punishment has two sides: it can be both loved and feared (Gordon et al., 2014; Kroupa, 2014). Previous studies have found that punishers can win some good reputation, which, however, may not necessarily translate into practical benefits (Ozono and Watabe, 2012).In addition, the effect of punishment on promoting cooperation has also been questioned. Mulder et al. (2006) had found that when the third-party punishment was added, the cooperation of players in PGG would increase, but when the punishment was withdrawn, the cooperation would become less. Other studies have made similar observations (Wang and Chen, 2011; Cui et al., 2017). Sometimes, the violator even bribe the third-party punishers to avoid being punished (Liu L. et al., 2019). Once corruption occurs, the third-party punishers will not only fail to promote cooperation but will undermine it.

Since the effect of third-party punishment in realizing indirect reciprocity and promoting cooperation is uncertain, what is its significance? For third-party punishment, we should distinguish between the existence and use of it. The existence of a third-party punishment mechanism can provide deterrence to group members and reduce their violations to promote cooperation, but the implementation of it may reduce cooperation. Therefore, a third-party punishment mechanism could exist, but the third party should use it less (Chen et al., 2014; Kroupa, 2014). Even if punishment is imposed, the mild punishment is more appropriate than the intensive punishment (Chen et al., 2015). This is akin to police armed with guns, who can deter criminals, but rarely fire; even when firing, such guns are usually limited in power and are just used to stop crimes but not to kill criminals. In addition, a more reasonable third-party punishment mechanism should be adopted. For example, the use of collective punishment rather than individual punishment can avoid being retaliated and reduce the cost of punishment (Sigmund et al., 2011). Besides, sharing punishment responsibility based on probability can also reduce costs (Chen et al., 2014). The reduction in cost is beneficial to the sustain of punishment (Wang et al., 2020).

Although one's reputation can be negative, people's real preference may be punishment. In Study 2b, we found that the third parties tended to choose punishment under the secret context but tended to choose compensation under the open context, indicating that the real reaction of third parties to unfair events is punishment, while compensation is just a strategic response to win a good reputation. Liu Y. et al. (2019) had found that under time pressure, the priority response of third parties is punishment; thus, they argued that people's instinctive response to unfair events is anger and punishment. In addition, in another study by Liu et al. (2017), it was found that third parties tend to punish in a "gain" context but compensate in a "loss" context. The present study was conducted under the gain context; hence, under the secret context, the real reaction of the third parties may be punishment.

Limitation

In the present study, the third parties either chose between keep and punish/compensate or between keep, punish, and compensate, which was a single choice; that is, they could only respond to unfair events in one way. In fact, the third parties could have another choice, in what is known as "double choice," which means that the third parties could choose punishment and compensation at the same time. A double choice includes both anger at the violator and empathy for the victim; hence, the motivation behind such a choice would be more complicated. Thus, this raises a series of questions: How do fourth parties understand the motivation of double election, and what reputation evaluation would be given to this motivation? Do third parties have a preference between single choice and double choice? To answer these questions, double choices should be considered in future research.

CONCLUSION

The difference in reputations of the third-party punishment and compensation is modulated by whether one is free to choose between the two methods. Specifically, when a third party is not free to choose, there is no reputational difference between the two methods, and the reputation of the two methods are all better than that of non-intervention. However, when there is the freedom to choose, the reputation of compensation is better than that of punishment, and the reputation of the two methods is also better than that of non-intervention.

To gain a good reputation, a third party will conduct reputational management through the choice between punishment and compensation, and the choice should also be mediated by whether there is the freedom to choose as well as the reputational context. Specifically, when there is no freedom to choose between the two methods, punishment and compensation under the open context occur more compared with the secret context, and there is no preference for the third party between the two methods. When there is the freedom to choose between the two methods, third-party punishment and compensation under the open context also occur more compared with the secret context; third parties prefer punishment under the secret context, while prefer compensation under the open context.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: The datasets for this study can be found in the "figshare." URL: https://doi.org/10.6084/m9.figshare.14139971.v1.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Academic Ethics Committee of Shanghai Dianji University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

ZL conceived the paper, ran statistical analyses, and contributed to the manuscript. GH conceived the paper and contributed to the manuscript. LX conducted the experiments. QL contributed to the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Shanghai Philosophy and Social Science Planning Project: Cognitive Neural Mechanism of Altruistic Behavior: Based on the Perspective of Time Pressure and Moral Sense (2020BSH007) and the Outstanding Clinical Discipline Project of Shanghai Pudong (PWYgy2018-10).

REFERENCES

Alexander, R. D. (1986). Ostracism and indirect reciprocity: the reproductive significance of humor. *Ethol. Sociobiol. 7*, 253–270. doi: 10.1016/0162-3095(86)90052-X

Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evol. Hum. Behav.* 27, 325–344. doi: 10.1016/j.evolhumbehav.2006.01.003

Bendor, J., and Swistak, P. (2001). The evolution of norms. *Am. J. Sociol.* 106, 1493–1545. doi: 10.1086/321298

Bereczkei, T., Birkas, B., and Kerekes, Z. (2007). Public charity offers as a proximate factor of evolved reputation-building strategy: an experimental analysis of a real-life situation. *Evol. Human Behav.* 28, 277–284. doi: 10.1016/j.evolhumbehav.2007.04.002

Burnham, T. C. (2003). Engineering altruism: a theoretical and experimental investigation of anonymity and gift giving. *J. Econ. Behav. Organ.* 50, 133–144. doi: 10.1016/S0167-2681(02)00044-6

Capraro, V., Jordan, J. J., and Tappin, B. M. (2021). Does observability amplify sensitivity to moral frames? Evaluating a reputation-based account of moral preferences. *J. Exp. Soc. Psychol.* 94:104103. doi: 10.1016/j.jesp.2021.104103

Capraro, V., and Perc, M. (2021). Mathematical foundations of moral preferences. *J. R. Soc. Interface* 18:20200880. doi: 10.1098/rsif.2020.0880

Capraro, V., and Rand, D. G. (2018). Do the right thing: experimental evidence that preferences for moral behavior, rather than equity or efficiency *per se*, drive human pro-sociality. *Judgment Decis. Mak.* 13, 99–111. doi: 10.2139/ssrn.2965067

Chavez, A. K., and Bicchieri, C. (2013). Third-party sanctioning and compensation behavior: findings from the ultimatum game. *J. Econ. Psychol.* 39, 268–277. doi: 10.1016/j.joep.2013.09.004

Chen, X., Szolnoki, A., and Perc, M. (2014). Probabilistic sharing solves the problem of costly punishment. *New J. Phys.* 16:083016. doi: 10.1088/1367-2630/16/8/083016

Chen, X., Szolnoki, A., and Perc, M. (2015). Competition and cooperation among different punishing strategies in the spatial public goods game. *Phys. Rev. E* 92:012819. doi: 10.1103/PhysRevE.92.012819

Condon, P., and Desteno, D. (2011). Compassion for one reduces punishment for another. *J. Exp. Soc. Psychol.* 47, 698–701. doi: 10.1016/j.jesp.2010.11.016

Cui, L., He, X., Luo, J., Huang, X., Cao, J., and Chen, X. (2017).The effects of moral punishment and relationship punishment on junior middle school students' cooperation behaviors in public goods dilemma. *Acta Psychol. Sin.* 49, 1322–1333. doi: 10.3724/SP.J.1041.2017.01322

Fehr, E., and Fischbacher, U. (2003). The nature of human altruism. *Nature* 425, 785–791. doi: 10.1038/nature02043

Fehr, E., and Fischbacher, U. (2004). Third-party punishment and social norms. *Evol. Hum. Behav.* 25, 63–87. doi: 10.1016/S1090-5138(04)00005-4

Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature* 415, 137–140. doi: 10.1038/415137a

Gordon, D. S., Madden, J. R., and Lea, S. E. G. (2014). Both loved and feared: third party punishers are viewed as formidable and likeable, but these reputational benefits may only be open to dominant individuals. *PLoS ONE* 9:e110045. doi: 10.1371/journal.pone.0110045

Horita, Y. (2010). Punishers may be chosen as providers but not as recipients. *Lett. Evol. Behav. Sci.* 1, 6–9. doi: 10.5178/lebs.2010.2

Horita, Y., and Takezawa, M. (2014). Observation enhances third-party punishment only among people who were not hot-tempered. *Lett. Evol. Behav. Sci.* 5, 5–8. doi: 10.5178/lebs.2014.28

Hu, Y., Sabrina, S., and Weber, B. (2015). Helping or punishing strangers: neural correlates of altruistic decisions as third-party and of its relation to empathic concern. *Front. Behav. Neurosci.* 9:24. doi: 10.3389/fnbeh.2015.00024

Jordan, J. J., Hoffman, M., Bloom, P., and Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature* 530, 473–476. doi: 10.1038/nature16981

Jordan, J. J., and Rand, D, G. (2020). Signaling when no one is watching: a reputation heuristics account of outrage and punishment in one-shot anonymous interactions. *J. Pers. Soc. Psychol.* 118, 57–88. doi: 10.1037/pspi0000186

Kroupa, S. (2014). Love or fear: can punishment promote cooperation? *Evol. Anthropol.* 23, 229–240. doi: 10.1002/evan.21430

Kurzban, R., DeScioli, P., and O'Brien, E. (2007). Audience effects on moralistic punishment. *Evol. Hum. Behav.* 28, 75–84. doi: 10.1016/j.evolhumbehav.2006.06.001

Leliveld, M. C., Van Dijk, E., and van Beest, I. (2012). Punishing and compensating others at your own expense: the role of empathic concern on reactions to distributive injustice. *Eur. J. Soc. Psychol.* 42, 135–140. doi: 10.1002/ejsp.872

Liu, L., Chen, X., and Szolnoki, A. (2019). Evolutionary dynamics of cooperation in a population with probabilistic corrupt enforcers and violators. *Math. Models Methods Appl. Sci.* 29, 2127–2149. doi: 10.1142/S0218202519500428

Liu, Y., Li, L., Zheng, L., and Guo, X. (2017). Punish the perpetrator or compensate the victim? Gain vs. loss context modulate third-Party altruistic behaviors. *Front. Psychol.* 8:2066. doi: 10.3389/fpsyg.2017.02066

Liu, Y., Wang, H., Li, L., Wang, Y., Peng, J., and Baxter, D. F. (2019). Judgments in a hurry: time pressure affects how judges assess unfairly shared losses and unfairly shared gains. *Scand. J. Psychol.* 60, 203–212. doi: 10.1111/sjop.12532

Lotz, S., Baumert, A., Schlösser, T., Gresser, F., and Fetchenhauer, D. (2011). Individual differences in third-party interventions: how justice sensitivity shapes altruistic punishment. *Negot. Confl. Manage. Res.* 4, 297–313. doi: 10.1111/j.1750-4716.2011.00084.x

Mulder, L. B., Dijk, E. V., Cremer, D. D., and Wilke, H. A. M. (2006). Undermining trust and cooperation: the paradox of sanctioning systems in social dilemmas. *J. Exp. Soc. Psychol.* 42, 147–162. doi: 10.1016/j.jesp.2005.03.002

Ozono, H., and Watabe, M. (2012). Reputational benefit of punishment: comparison among the punisher, rewarder, and non-sanctioner. *Lett. Evol. Behav. Sci.* 3, 21–24. doi: 10.5178/lebs.2012.22

Piazza, J., and Bering, J. M. (2008). The effects of perceived anonymity on altruistic punishment. *Evol. Psychol.* 6, 487–501. doi: 10.1177/147470490800600314

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

Raihani, N. J., and Bshary, R. (2015b). The reputation of punishers. *Trends Ecol. Evol.* 30, 98–103. doi: 10.1016/j.tree.2014.12.003

Rockenbach, B., and Milinski, M. (2011). To qualify as a social partner, humans hide severe punishment, although their observed cooperativeness is decisive. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18307–18312. doi: 10.1073/pnas.1108996108

Rodrigues, J., Nagowski, N., Mussel, P., and Hewig, J. (2018). Altruistic punishment is connected to trait anger, not trait altruism, if compensation is available. *Heliyon* 4:e00962. doi: 10.1016/j.heliyon.2018.e00962

Santos, M. D., Rankin, D. J., and Wedekind, C. (2013). Human cooperation based on punishment reputation. *Evolution* 67, 2446–2450. doi: 10.1111/evo.12108

Schram, A., and Charness, G. (2015). Inducing social norms in laboratory allocation choices. *Manage. Sci.* 61, 1531–1546. doi: 10.1287/mnsc.2014.2073

Sierksma, J., Thijs, J., and Verkuyten, M. (2014). Children's intergroup helping: the role of empathy and peer group norms. *J. Exp. Child Psychol.* 126, 369–383. doi: 10.1016/j.jecp.2014.06.002

Sigmund, K., Hauert, C., Traulsen, A., and Silva, H. (2011). Social control and the social contract: the emergence of sanctioning systems for collective action. *Dyn. Games Appl.* 1, 149–171. doi: 10.1007/s13235-010-0001-4

Wang, P., and Chen, L. (2011). The effects of sanction and social value orientation on trust and cooperation in public goods dilemmas. *Acta Psychol. Sin.* 43, 52–64. doi: 10.3724/SP.J.1041.2011.00052

Wang, Q., Liu, L., and Chen, X. (2020). Evolutionary dynamics of cooperation in the public goods game with individual disguise and peer punishment. *Dyn. Games Appl.* 10, 764–782. doi: 10.1007/s13235-019-00339-7

Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S., and Cook, K. S. (2009). The private rejection of unfair offers and emotional commitment. *Proc. Natl. Acad. Sci. U.S.A* 106, 11520–11523. doi: 10.1073/pnas.0900636106

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.