Aligning with the Agent of Justice:

Schadenfreude Following Punishment of Trust Violators

Paton Pak Chun Yam

Department of Psychology, De Montfort University

Fan Huang

Institute of Logic and Cognition, Sun Yat-Sen University

Xudong Luo

School of Computer Science and Engineering, Guangxi Normal University

Brian Parkinson

Department of Experimental Psychology, University of Oxford

Author Note

These studies were conducted while the second author was a visiting student at Oxford University, UK. The final author's contribution to this research was supported by the Economic and Social Research Council (UK, grant numbers: RES-060-25-0044 and ES/L016486/1).

Please address all correspondence to the first author, Paton Pak Chun Yam at the Department of Psychology, De Montfort University, The Gateway, Leicester, LE1 9BH, United Kingdom. E-mail: paton.yam@dmu.ac.uk

Conflict of Interest Statement

The authors have no conflict of interest to declare. The authors also confirm that the research reported in this paper adheres to ethical guidelines specified in the APA Code of Conduct as well as authors' national ethics guidelines.

Data Availability Statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Abstract

Four experiments used trust games to investigate schadenfreude's effects on attitudes and behavior towards third parties who punish prior violations of participants' trust. Across all studies, schadenfreude was stronger when trust violators received negative rather than positive outcomes, and participants' perceptions that the violator deserved punishment positively predicted levels of schadenfreude. Further, participants had less favorable attitudes towards third parties who delivered more negative outcomes (Experiment 1), but attitudes and behavior were relatively more favorable when these negative outcomes were inflicted on trust violators (Experiment 2). Participants also had more favorable attitudes and behavior towards third parties who delivered punitive outcomes which did not themselves involve trust violations. Further, results were consistent with a statistical model where just world beliefs, deservingness, and schadenfreude serially mediated the effect of punishment on attitudes towards third-party punishers (Experiment 3). These results were replicated in the pre-registered Experiment 4 using a larger sample and more focused measures. Taken together, our findings are consistent with the view that schadenfreude following trust violations serves the social function of aligning relations with agents who administer retributive justice.

Key words: Deservingness, interpersonal relations, justice, schadenfreude, trust games

Aligning with the Agent of Justice:
Schadenfreude Following Punishment of Trust Violators

When other people get punished after causing us to suffer, we sometimes take secret pleasure in their misfortune – an emotional response best captured by the German word "schadenfreude" (van Dijk & Ouwerkerk, 2014). These four experiments investigated schadenfreude in response to punishment of trust violators, and its consequences for attitudes and behavior towards the person delivering punishment.

Deservingness and Justice

According to deservingness theory, appraising other people's positive and negative outcomes as deserved or undeserved leads to different emotional outcomes (Feather, 2006). For example, we experience schadenfreude when we appraise negative outcomes as deserved, but sympathy when we appraise negative outcomes as undeserved. Research has also shown that the perceived deservingness of a misfortune positively predicts schadenfreude (Feather & Sherman, 2002; van Dijk et al., 2008) and mediates effects of the victim's appraised personal responsibility on this emotion (Berndsen & Feather, 2016; van Dijk et al., 2015).

These findings suggest that justice concerns may underlie many instances of schadenfreude (Wang et al., 2019). As Portman (2002) noted: "It is not the suffering of others that brings us joy, but rather the evidence of justice triumphing before our eyes" (p. xiii). In other words, people should feel schadenfreude when a norm-violator receives a deserved punishment, thus correcting the prior norm violation and restoring justice. Indirect evidence for this prediction comes from studies showing that people feel stronger schadenfreude after reading stories about a disliked classmate receiving a deserved penalty for plagiarism (Feather & Mckee, 2014) or that participants spent longer reading stories about schadenfreude when their belief in a just world had been threatened (Pietraszkiewicz, 2013). The present research extends these findings by focusing on consequences rather than

antecedents in order to clarify the social functions of schadenfreude in justice-related contexts.

Social Functions

Emotions serve important social functions (Keltner & Haidt, 1999). At the dyadic level, they communicate appraisals and intentions to other people (Parkinson et al., 2005) and serve both social affiliation and distancing functions (Fischer & Manstead, 2016) . A growing body of research has shown that schadenfreude may serve a social distancing function involving comparison and competition. For instance, schadenfreude has been associated with negative word of mouth (Sundie et al., 2009), willingness to harm (Cikara et al., 2011), withholding helping behavior (Schindler et al., 2015), and cheating (Lange & Boecker, 2019). These actions effectively create social boundaries between the emoter and the target of schadenfreude, allowing the emoter to engage in downward comparison resulting in the attainment of higher social status or more power (Fischer & Manstead, 2016).

Unlike these previous studies, we aim to add to the growing body of evidence that schadenfreude may serve a social affiliative function (e.g. Brambilla & Riva, 2017). We argue that schadenfreude in justice-related contexts serves to align social relations by making people more positively disposed towards third parties who administer deserved punishment, thereby upholding justice in our social relationships. In support of this claim, indirect evidence shows that participants rate third-party punishers more positively when they intervene following norm-violating behaviors (Gordon et al., 2014), reward them when they punish the dictator in dictator games (Raihani & Bshary, 2015), and prefer them as partners when they inflict greater costs on trust violators (Nelissen, 2008). However, other studies suggest that participants do not always reward third-party punishers (Kiyonari & Barclay, 2008). This discrepancy may depend on whether or not participants experience schadenfreude in response to observed punishment. In the present study, we predict that participants who

feel schadenfreude when a third party punishes a trust violator should be more positively disposed towards that third party because of their satisfaction that justice has been restored.

Trust Violation

Instead of responding to vignettes describing imagined justice-related outcomes as has been common in previous schadenfreude research (e.g. Feather & Mckee, 2014; Feather & Sherman, 2002), our participants played interpersonal trust games (Berg et al., 1995), where they experienced directly involving trust violations and witnessed the trust violator's subsequent outcomes in real time. We focused on trust violations because of the importance of trust in developing and maintaining interpersonal relationships (Rousseau et al., 1998). Trust violations and other people's responses to them therefore provide an engaging and meaningful context for assessing the social function of schadenfreude.

Trust has been defined as the willingness to accept a vulnerable relation with others, built upon a positive expectation of their intentions or behaviors (Rousseau et al., 1998). When trust is violated, victims may have a sense of injustice because the trust violator has failed to meet their positive expectations (Bies & Tripp, 1996). Consequently, victims of trust violations are motivated to punish trust violators and take revenge (Fehr & Gächter, 2002) because these actions can help restore justice (Darley & Pittman, 2003).

Third-Party Punishment

In many cases, the victim is not the only person capable of restoring justice following a trust violation. For example, third-party observers readily punish violators who fail to reciprocate another person's trust (Charness et al., 2008), especially when they evaluate the violation in negative terms (Fehr & Fischbacher, 2004). Punishment by either the victim or a third party serves to deter future norm violations and deliver retribution (Crockett et al., 2014).

Insofar as the wrongdoer receives punishment that is perceived as proportionate

(Carlsmith et al., 2002) and deserved (Feather, 2006), third-party punishment should satisfy victims' need for justice and consequently lead to schadenfreude. Indeed, previous research has shown that people feel more positive when they witness an offender being punished by a third party than when they instigate revenge against that person themselves (Carlsmith et al., 2008). We therefore predicted that third-party punishment of trust violations would induce schadenfreude.

The Present Research

In sum, the present research had two aims: first, to determine whether schadenfreude is experienced following punishment of trust violations; and second, to assess schadenfreude's effects on attitudes and behavior towards third-party punishers. In all four experiments, participants took the role of investor in an initial trust game and received a trust-violating return from the other player. They then observed that the other player either receive or not receive a negative outcome from a third party.

The four experiments reported in this paper examined the boundary conditions and underlying mechanisms behind the affiliative function of schadenfreude towards the third party in justice-related contexts. Experiment 1 compared attitudes and behavior towards third parties who delivered either a negative or positive outcome towards the trust violator. Experiment 2 compared attitudes and behavior towards third parties who delivered negative or positive outcome either to the trust violator or another player. Experiment 3 compared attitudes towards punishers and non-punishers when punishment did not itself involve a further trust violation. Experiment 4 was a preregistered study intended to replicate the main findings of Experiment 3. Across all four studies, we predicted that the third party's delivery of negative outcomes to the trust violator would increase participants' schadenfreude and improve their attitudes and behavior towards that third party.

Experiment 1

Experiment 1 assessed whether observing a player who received a deserved negative outcome after violating the participant's trust leads both to schadenfreude and to more favorable attitudes and behavior towards the third party who was responsible for delivering this outcome. We independently manipulated whether the trust violator received a positive or negative outcome (punishment manipulation) and whether the participant played a subsequent trust game with the third party or a different player (final investor manipulation).

The punishment manipulation allowed direct comparison between emotional reactions to the third party's delivery of deserved and undeserved outcomes to the trust violator. Based on Feather (2006)'s model of deservingness, the judgement of deservingness depends on the relation between action and outcome. Positive/negative outcomes produced by positive/negative actions are perceived as deserved. In contrast, positive/negative outcomes that follow from negative/positive actions are seen as undeserved. Hence, punishment following a trust violation should be seen as more deserved than no punishment and should consequently increase schadenfreude to a greater extent, and this in turn should result in improvement in participants' attitudes and behavior towards the third-party punisher. In other words, we hypothesized that deservingness and schadenfreude would sequentially mediate the effect of punishment on attitudes and behavior towards the third-party punisher.

The final investor manipulation was designed to assess whether punishment specifically improved interpersonal behavior directed towards the punisher rather than having more general positive effects on social relations. Reflecting schadenfreude's proposed function of aligning with the agent of justice, we predicted that participants would behave more favorably only towards the third party who had previously punished the trust violator and not to another player who had no prior dealings with the trust violator.

Method

Participants

Power analysis using the program G*Power (Faul et al., 2007) determined that 90 participants were required to detect a medium effect size (Cohen, 1988) with a power of .80 in a two-way ANOVA of the interactive effects of Punishment and Final investor manipulations on behavior. We recruited 93 participants (63 women; M_{age} = 21.00, *SD* = 2.77) who were students or visiting students at the University of Oxford, UK. Twelve were excluded from analysis because they reported suspicions about whether they were genuinely playing against other human players during funnelled debriefing (Bargh & Chartrand, 2000) which made it possible that the emotions reported by these participants were not genuinely felt.

Design

The experiment used a 2 (Punishment: Punishment vs. No-punishment) × 2 (Final investor: Third party vs. Different investor) between-subjects design, and participants were randomly allocated to the resulting four conditions. Because the Final Investor manipulation was not implemented until the final game, analyses of prior dependent measures used independent samples t-tests assessing effects of the Punishment manipulation only. The main dependent variables were deservingness, schadenfreude, and attitudes and behavior towards the third party[1].

Procedure

Participants arrived at the laboratory in pairs and were told that two other participants were also due to arrive. They were then taken to individual cubicles to play the games on-line. Each participant in fact played separately with simulated other players whose responses were pre-programmed using Qualtrics software (Provo, UT, 2015).

The experimenter explained the rules of the trust game and told participants that the

investor would be provided with an initial stake of 100 points in each game, and that both players should try to earn at least 110 points in each game in order to win a chocolate bar, which was placed prominently on the table in the room where the experiment took place. The experimenter further advised participants that any points the investor transferred would be multiplied by 4 before reaching the trustee and clarified that sending 40 points would provide sufficient points for both players to win the chocolate bar (on the assumption that the trustee sent back 50 of the resulting 160 (40 points x 4) transferred points). Participants also learnt that the two participants not taking the role of either investor or trustee in each game would monitor its progress on-line as "observers."

Participants took the role of investor in game 1. If they did not initially transfer at least 40 of their 100 points, a pre-programmed prompt reminded them of this advised transfer amount.[2] After transferring the chosen number of points, participants saw a screen showing that the trustee had returned zero points, thus violating their trust.

Participants then observed game 2, in which the trust violator from game 1 transferred 50 points as investor and asked for 75 points in return. The game 2 trustee's behavior constituted the punishment manipulation. In the *punishment* condition, the trustee returned zero points (meaning that only the trustee was left with sufficient points to win a chocolate bar). In the *no punishment* condition, the trustee returned 75 points (meaning that both players earned enough points to win a chocolate bar). After observing the outcome of this second game, participants completed an on-line questionnaire about schadenfreude, their perception of how much the trust violator deserved the outcome, and their attitudes towards the third party.

Participants were assigned the role of trustee in the third game (which they did not yet know was the final game). The identity of the investor in this game constituted the final investor manipulation. In the *third party* condition, the investor was the same player who had

delivered the punishing or non-punishing outcome in game 2; in the *different investor* condition, the investor had observed games 1 and 2 but not played until this game. Regardless of player identity, participants received 160 points (40 X 4) from the game 3 investor, who asked them to return 50 points. After choosing how many points to return, participants were debriefed.

Measures

Schadenfreude and deservingness. To assess schadenfreude following game 2, we combined two items adapted from van Dijk et al.'s (2008) scale ("How satisfied are you with what happened to [the target]?" and "I couldn't resist a little smile when I think of what happened to [the target]?", 1 = "*Strongly disagree*", 7 = "*Strongly agree*"), $r(79) = .44$, $p < .001$. Participants rated deservingness on a scale from 1 ("*Not at all*") to 7 ("*Very much*") using a single item assessing how much the trust violator deserved the outcome of game 2.

Attitudes. We measured participants' attitudes towards the third party following game 2 with two items ("How much do you like [the target]?", and "How highly do you regard [the target]?", 1 = "*Not at all*", 7 = "*Very much*"). These items showed a strong positive correlation, $r(79) = .66$, $p < .001$, and were combined into a composite score.

Behavior. The number of points returned to the investor in the final game provided our measure of behavior towards the third party.

Results

Schadenfreude and deservingness

As predicted, both schadenfreude and deservingness were significantly higher in the punishment condition (schadenfreude: $M = 4.46$, $SD = 1.58$; deservingness: $M = 4.53$, $SD = 1.97$) than in the no-punishment condition (schadenfreude $M = 3.08$, $SD = 1.21$, $t(79) = 4.43$, $p < .001$, $d = 0.98$; deservingness $M = 3.45$, $SD = 1.62$, $t(78) = 2.67$, $p = .009$, $d = 0.56$)[3].

Attitudes

Contrary to prediction, attitudes towards the third party were significantly more positive in the no punishment condition (M = 5.26, SD = .99) than in the punishment condition (M = 3.77, SD = 1.27), $t(79) = 5.92$, $p < .001$, $d = 1.31$.

Returned points

Most participants (87.65%) returned the requested 50 points to the investor in the final game regardless of condition. There was no significant effect of the punishment or final sender manipulation.

Correlations

In the punishment condition, schadenfreude showed reliable positive correlations with deservingness, $r(39) = .51$, $p = .001$, and attitudes to the third-party punisher, $r(39) = .51$, $p = .001$. Deservingness was also positively correlated with attitudes towards the third party, $r(38) = .42$, $p = .007$.

Mediation analysis

Because the effect of punishment on attitudes was opposite to our prediction, we did not test the full sequential mediation model including that outcome variable and instead only assessed the indirect effect of punishment on schadenfreude via deservingness using Hayes'(2022) PROCESS macro (Model 4) with 5000 bootstrap samples to obtain 95% bias-corrected confidence intervals to explore whether the data supported previous evidence for the role of deservingness in schadenfreude (Feather, 2014). Results revealed that the total effect (b = 1.36, SE = .32, t = 4.30, $p < .001$) of punishment on schadenfreude was significantly reduced (b = .89, SE = .28, t = 3.23, p = .002) after controlling for the effects of deservingness (indirect effect = .47, SE = .20, 95% CI = [.121 — .910]).

Discussion

As expected, participants reported higher levels of schadenfreude after game 2 and believed that the trust violator deserved the received outcome to a greater extent in the punishment than in the no-punishment condition. Deservingness also correlated positively with schadenfreude in the punishment condition and controlling for its effects significantly reduced the effect of the punishment manipulation on schadenfreude, consistent with previous research on the mediational role of this variable (e.g. Feather & Mckee, 2014; Feather & Sherman, 2002).

Contrary to prediction, participants had relatively more favorable attitudes towards third parties in the no-punishment condition. This probably reflected their perception that the delivery of punishment itself constituted a further trust violation leading to negative reactions that outweighed any positive effect of deservingness. Experiment 2 therefore used a different comparison condition that allowed us to control for any negative impact of the punisher's violation of the investor's trust.

Despite these potential issues with Experiment 1's punishment manipulation, schadenfreude was positively correlated with deservingness and attitudes towards the third party in the punishment condition. Thus, participants experiencing stronger schadenfreude after witnessing delivery of a deserved negative outcome were related to having relatively more positive feelings about the agent of justice. Taken together, these findings were consistent both with our general predictions and with previous findings supporting the role of deservingness in schadenfreude (Feather & Sherman, 2002).

We were unable to detect effects on our intended behavioral dependent measure in Experiment 1 because a substantial majority of participants returned exactly 50 points to the investor in the final game. Two factors potentially accounted for this general lack of variance. First, fairness norms may have influenced behavior more than we had anticipated,

encouraging participants to accede to the other player's request for an equitable return of 50 points. Indeed, fairness may have been particularly salient because participants met each other before the experiment, thus increasing their sense of identifiability. Second, there was no practical disincentive against returning the requested 50 points to the other player in the final game because the remaining 110 points were still sufficient to earn the reward of the chocolate bar. Experiment 2 attempted to deal with these two issues.

Experiment 2

Experiment 2 aimed to clarify Experiment 1's findings by comparing a condition in which the trust violator was punished with one in which an equally negative outcome was delivered to a player who had not committed any trust violation. This modification allowed us to focus on specific effects of restoring justice after factoring out any negative impact of the associated trust violation. We predicted that participants would be relatively more positively disposed towards a third party who had punished the trust violator than towards someone who had delivered an equally negative outcome to another player who had committed no prior trust violation and therefore did not deserve to be punished. Consequently, deservingness and schadenfreude should mediate the effect of the punishment manipulation on attitudes towards the third party after removing the negative impact of the trust violation on attitude.

Experiment 2 also attempted to minimise the impact of factors restricting variance in the number of points returned in the final game. We reduced the salience of fairness norms by removing the possibility of meeting other players before the game. Potential participants were emailed with on-line links to the study which they completed in private and with total and evident anonymity. We also removed the incentive of a chocolate bar so that participants would strive to maximise their overall points in return for lottery opportunities rather than

only aiming to achieve a predefined points threshold in each game.

Method

Participants

We attempted to recruit a sample as large as we used in experiment 1, choosing to keep the study link open for a sufficiently long period to reach the target and with the expectation that several participants would not provide usable data (e.g., because of failed attention checks or withdrawing participation). Unexpectedly, only 64 participants (35 women; M_{age} = 22.08, *SD* = 4.83) responded to our initial email advertising the study at the University of Oxford, UK[4]. We therefore supplemented the sample with additional students from the Chinese University of Hong Kong using the university's mass-mailing service. This additional data collection yielded more participants. Fifty-one (20%) participants failed to complete the full procedure for various reasons[5] and were excluded from analyses. The final sample included 109 women and 56 men (*N* = 165; M_{age} = 20.64, *SD* = 3.39). No analyses were conducted prior to the conclusion of this data collection.

Design

The experiment again used a 2 (Punishment: Punishment *vs.* No punishment) × 2 (Final investor: Third party *vs.* Different investor) between-subjects design, with the final investor manipulation only affecting behavioral data collected in the final game. Our inclusion of an additional comparison game (2b, see below) where outcomes were delivered to a virtual player other than the trust violator also introduced an additional within-subjects variable of Punishment target (Trust violator *vs.* New target).

Procedure

The procedure was similar to Experiment 1, except for the inclusion of additional virtual players and one further trust game to preserve equivalence of all relevant variables

across all conditions (see below). Participants followed a worked example of a trust game, answered comprehension questions about its rules, and received immediate feedback indicating correct responses before completing a practice game themselves. These modifications were intended to ensure that participants understood how the games worked in Experiment 2's on-line setting, which permitted no direct clarification from a responsive experimenter (unlike in Experiment 1).

Participants were informed that five players would be involved in the subsequent games, and that any points they earned would increase their chances of winning a £50 (~USD $65) prize in a lottery. Introducing this more direct relation between points and reward was intended to encourage participants to maximize their total score in the final game rather than simply aiming to reach a predefined threshold as in Experiment 1, thus making the transfer of points to the other player a more meaningful measure of their positive intentions. Further, ostensibly to preserve anonymity, but in fact to improve clarity and memory for experienced and witnessed outcomes, each player was allocated a randomly selected animal name (Sheep, Rabbit, Dolphin, Gorilla, or Horse) and color-coded animal icon. See Table 1 for an outline of the structure of the games.

T1. The participant took the role of investor in game 1. The trustee was identified as Rabbit, and there were two observer participants (one of whom would deliver the punishing or non-punishing outcome to Rabbit in Game 2a, and the other of whom would deliver an equally negative or positive outcome to a new player in Game 2b). The participant was given an initial allocation of 100 points and chose how many of these points to send to Rabbit. Rabbit returned only 2 points, thus violating the participant's trust.

T2. Participants were assigned the role of (sole) observer in the second and third games (2a and 2b), whose order was counterbalanced. Game 2a followed a similar pattern to game 2 in experiment 1 and involved one of game 1's observers taking the role of trustee and

either punishing or not punishing game 1's trust violator, Rabbit, who acted as the investor in game 2a.

The investor in Game 2b was a new player, Gorilla, who received comparable outcomes to those delivered to Rabbit in Game 2a. Thus, this new player received a negative outcome from the trustee in the punishment condition and a positive outcome from the trustee in the no-punishment condition. The inclusion of this additional game thus permitted us to directly compare participants' reactions to punishment (or non-punishment) in game 2a with equivalent outcomes that did not involve punishment (or non-punishment) of a prior trust violation in game 2b.

T3. The investor in the final game was the player who had been the trustee in either game 2a (Third party) or game 2b (Different investor) depending on the Final investor manipulation, and the trustee was the participant. Both investors had previously delivered identical outcomes to a trustee (either positive or negative depending on punishment condition), but only the *Third party* had delivered this outcome to the initial trust violator (Rabbit); the *Different investor* had delivered the outcome to the other player (Gorilla). This use of different investors permitted us to make direct comparisons between participants' responses after controlling for negativity of outcomes (since identical outcomes were delivered by both investors, but the outcome could only be perceived as a punishment in the Third party condition). The T3 investor transferred 60 points to the participant (multiplied by 4 to yield 240 points) and asked for 120 points to be returned. After choosing how many points to return, participants completed a final questionnaire about their perceptions of the T3 investor before debriefing.

Measures

All measures were identical to those used in Experiment 1 with the following exceptions. Because of potential concerns about the reliability of the two-item schadenfreude

scale used in Experiment 1, we added two more of van Dijk et al.'s (2008) items ("How much did you enjoy what happened to [the target]?" and "I actually had to laugh a bit when I think of what happened to [the target]") to this measure. The resulting 4-item scale, which was administered after game 2a, showed a satisfactory level of reliability (α = .83). Reliability of the 2-item scale assessing participants' attitudes towards the trustee, $r(163) = .90$, $p < .001$, and towards the other player at T2, $r(163) = .96$, $p < .001$,were high, allowing for the computation of a difference score.)

Results

Schadenfreude and deservingness

As predicted, both schadenfreude ($t(138.88) = 2.86$, $p = .005$, $d = 1.38$) and deservingness ($t(163) = 3.55$, $p = .001$, $d = 1.34$) after game 2a were significantly higher in the punishment condition (schadenfreude $M = 3.92$, $SD = 1.61$; deservingness $M = 4.59$, $SD = 1.94$) than in the no-punishment condition (schadenfreude $M = 3.29$, $SD = 1.13$; deservingness $M = 3.58$, $SD = 1.73$)[6].

Attitudes towards trustees at T2

A mixed design 2 (Punishment: Punishment *vs.* No-punishment) x 2 (Punishment Target: Trust violator *vs.* New target) ANOVA was conducted to compare the effects of delivering outcomes to the trust violator and to the new player. Results revealed a significant main effect of punishment on attitudes towards trustees, $F(1, 163) = 226.74$, $p < .001$, $\eta^2 = .582$, again showing less positive attitudes towards the T2 trustee in the punishment condition ($M = 2.66$, $SD = 1.12$) than in the no punishment condition ($M = 5.17$, $SD = 1.03$). However, this main effect was qualified by a significant Punishment by Punishment Target interaction, $F(1, 163) = 50.81$, $p < .001$, $\eta^2 = .238$. Simple main effects showed that attitudes towards the trustee in the punishment condition were significantly more positive when the

investor was the trust violator (M = 3.06, SD = 1.53) rather than the new target (M = 2.25, SD = 1.17), $F(1, 163) = 26.10$, $p < .001$, $\eta^2 = .138$. By contrast, attitudes in the no punishment condition were significantly more negative when the investor was the trust violator (M = 4.79, SD = 1.14) rather than the new target (M = 5.55, SD = 1.28), $F(1, 163) = 24.72$, $p < .001$, $\eta^2 = .132$. In other words, after factoring out the negativity of the delivered outcome, punishment of a trust violator led to more positive attitudes, whereas failure to punish led to more negative attitudes.

Returned points at T3

A 2 (Punishment: Punishment *vs.* No-punishment) × 2 (Final investor: Third Party *vs.* Different Investor) ANOVA revealed a main effect of punishment on the number of points returned by participants in the final game, $F(1, 161) = 17.16$, $p < .001$, $\eta^2 = .096$, with more points returned in the no-punishment condition (M = 100.29, SD = 44.66) than in the punishment condition (M = 69.10, SD = 54.06). This main effect was qualified by a significant punishment by final investor interaction, $F(1, 161) = 9.07$, $p = .003$, $\eta^2 = .053$. Simple main effects showed that participants in the punishment condition returned more points to an investor who had punished the trust violator (Third party: M = 86.20, SD = 54.67) than to an investor who had delivered an equally negative outcome to a different target (Different investor: M = 51.56, SD = 48.05), $F(1, 161) = 10.27$, $p = .002$, $\eta^2 = .060$. However, the corresponding difference was not significant in the no-punishment condition (Third party: M = 94.68, SD = 41.99; Different investor M = 105.17, SD = 46.77), $p = .314$.

Correlations

Similar to Experiment 1, schadenfreude towards the trust violator was reliably positively correlated with deservingness, $r(77) = .50$, $p < .001$, with attitudes towards the investor who punished the trust violator, $r(77) = .41$, $p < .001$, and with attitudinal difference between the third party and the different investor, $r(77) = .44$, $p < .001$, in the punishment

condition. Deservingness was again positively correlated with attitudes towards the punishing investor, $r(77) = .53$, $p < .001$. Only half of the participants in the punishment condition were paired with the punishing investor in the final game (i.e., those in the Third Party as Final investor condition). For this subset of participants, points returned in the final game showed no significant associations with attitudes, attitudinal difference between third party and the different investor, deservingness, or schadenfreude.

Serial mediation analyses

Unlike Study 1 in which the punishment manipulation was confounded with trust violation, Study 2 introduced the different investor condition to disentangle the effects of trust violation from punishment. Hence, we were able to obtain a purer measure of the effect of punishment on attitudes towards the third party by subtracting the attitudes towards the different investor (which was influenced by trust violation but not punishment). We used Hayes' (2022) PROCESS (Model 6) with 5000 bootstrap samples to test the serial indirect effect of punishment on the difference in attitudes via deservingness and schadenfreude. Results showed that the total ($b = 1.58$, $SE = .22$, $t = 7.13$, $p < .001$), direct ($b = 1.32$, $SE = .22$, $t = 6.04$, $p < .001$), and the total indirect effect were significant ($b = .26$, $SE = .11$, [.064 — .485]). Although the indirect effect of punishment on difference in attitudes through deservingness and schadenfreude were not significant, the serial indirect effect of deservingness and schadenfreude was significant (*indirect effect* = .11, $SE = .07$, 95% CI = [.023 — .271]), consistent with our hypothesis. *Figure* 1 shows all the paths for the serial mediation model.

Discussion

Experiment 2's findings extended those of Experiment 1. Participants again reported stronger schadenfreude and believed the trust violator's received outcome was more deserved

when that trust violator was punished. Further, procedural changes meant that participants' points allocations in the final game directly influenced their chances to enter the lottery, increasing the psychological realism of the experiment. Consequently, the punishment manipulation not only influenced their self-reported attitudes, but also their trust behaviors, which now showed sufficient variation to be responsive to our manipulations.

These changes allowed us to disentangle the interpersonal effects of punishment from those of delivering a negative outcome in the form of a trust violation (which were confounded in Experiment 1). Participants were again more positively disposed towards players who delivered positive outcomes than to players who delivered negative outcomes. However, as predicted, participants showed relatively more positive attitudes and returned more points to a player who had punished a prior trust violator than to a player who delivered a comparable negative outcome to someone who had not previously violated the participant's trust. The results of our analyses were also consistent with a statistical model where deservingness and schadenfreude serially mediated the effect of punishment on difference in attitudes towards the third party after removing the effect of delivering a comparable negative outcome to a non-trust violator.

Interestingly, participants also had more negative attitudes towards a player who failed to punish the prior trust violator than to a player who delivered a comparable positive outcome to a different player. In both cases, it seems that participants reacted more favorably to players who delivered outcomes that they perceived as more just and more deserved by the prior trust violator.

Experiment 3

Experiment 1's prediction was that punishment of trust violators would lead to more positive attitudes and behavior towards the punisher. However, negative outcomes generally

worsened interpersonal attitudes towards the person delivering them in both previous experiments, probably because these negative outcomes took the form of an additional trust violation by the third party punisher. The fact that the punisher violated the investor's trust (after the participant had experienced a similar trust violation) may have counteracted the positive effect of restoring justice. To address this issue, Experiment 3 assessed effects of punishment that did not involve trust violation, thereby removing any associated counter-normative implications. We predicted that delivery of a normative and deserved punishment would lead to stronger schadenfreude, and more positive attitudes and behavior towards the third party.

Experiment 3 also included a more direct test of our proposed explanation for the effects of punishment on deservingness and schadenfreude and on attitudes towards punishers. Experiment 1's findings were consistent with previous research showing that deservingness mediated the effect of punishment on schadenfreude (Feather & Mckee, 2014) and analyses of Experiment 2 data were consistent with a statistical model where deservingness and schadenfreude serially mediated the effect of punishment on attitudes towards the third-party punisher after removing the negative effects of trust violation on attitudes. Experiment 3 took the additional step of assessing belief in a just world (BJW, Lerner, 1980) in order to evaluate the possibility that punishment leads to schadenfreude because it helps to defend participants' BJW when it is threatened by experience of a trust violation.

According to Lerner's (1980) justice motive theory, people have a fundamental need to believe they live in a just world where people get what they deserve. This belief is beneficial because it makes one's surroundings more predictable and controllable. Hence, people are motivated to sustain this belief, which eventually leads them to construe and explain everyday experience around issues of deservingness. Further, in the face of injustice,

people are motivated to defend or restore their threatened BJW (for a review, see Hafer & Rubel, 2015). Seeing a perpetrator gets punished helps to mitigate such threat, reaffirming observers' belief that the world is just, where everyone gets what they deserve (Hafer, 2000). Indeed, past research has shown that stronger BJW is associated with quicker processing of deservingness-relevant than deservingness-irrelevant information (Hafer et al., 2020) and seeing negative situations as fairer or higher in deservingness (Olson & Hafer, 2001). More relevantly, a threat to BJW was found to increase the judgment of deservingness and encourage subsequent schadenfreude (Pietraszkiewicz, 2013).

Based on these previous findings, we argue that BJW's relation to deservingness and justice concerns may underlie schadenfreude's proposed social affiliative functions. Schadenfreude serves to align the person with agents who correct injustice in the interpersonal distribution of social outcomes (e.g. reward and punishment). In other words, it is an emotion that depends on the motivation to defend against potential threats to BJW (Hafer & Rubel, 2015). Following the results by Pietraszkiewicz (2013), it could therefore be argued that the experience and social consequences of schadenfreude depend on the sense that BJW has been restored which in turn leads to the judgment that the punitive process is deserved. Therefore, our predictions were that BJW would be higher after observing the trust violator being punished, and that higher BJW would positively predict deservingness, schadenfreude, and positive attitudes to the third-party punisher. In other words, observing punishment would affect attitudes to the third-party via a serial indirect effect of BJW, deservingness, schadenfreude.

Experiment 3 also included an additional comparison condition in which the initial trust violation was inflicted on a player other than the participant, to assess the possibility that retributive punishment has similar effects on schadenfreude and interpersonal attitudes even when the original trust violation is not directly self-relevant. In particular, we manipulated

self-relevance of the trust violation by comparing conditions in which the first game involved the participant (self-relevant condition) or a different player (non-self-relevant condition). We predicted that participants who had directly experienced a prior trust violation would experience stronger schadenfreude than those who had observed a trust violation inflicted on someone else. Further, we predicted that self-relevant trust violations would have relatively greater effects on interpersonal attitudes and behavior.

Method

Participants

The on-line study link for Experiment 3 was kept open for a period intended to yield a similarly sized sample to Experiment 2 after accounting for potential dropouts and exclusions. This resulted in the recruitment of 301 participants (using MTurk). All participants lived in USA and had an approval rate of at least 95% for their previous assignments. One hundred and forty-two participants either failed to complete the full procedure or produced problematic data and were excluded from analysis[7]. The final sample included 159 participants (83 women and 76 men; M_{age} = 33.42, SD = 10.44). No analyses were conducted until data collection was complete.

Design

Experiment 3 used a 2 (Punishment: Punishment *vs.* No-punishment) × 2 (Self-relevance: Self-relevant *vs*. Non-self-relevant) between-subjects design, and participants were randomly allocated to one of the four conditions. Our main dependent variables were BJW, deservingness, schadenfreude, attitudes towards other players, and points transferred after the second game.

Procedure

After accepting our job on MTurk, participants were linked to Qualtrics (Provo, UT,

2015). They were then taken through a worked example of a trust game and answered comprehension questions about its rules before learning the correct responses. This was to ensure their comprehension of the instructions and understanding of their decisions when no direct communication with the experimenter was possible. They then completed a practice trial, taking the role of investor. As in experiment 2, players were assigned animal names and colored icons and were endowed with 100 points before the trust games began.

Game 1 involved either the participant or another player taking the role of investor depending on the self-relevance manipulation. In the self-relevant condition, participants were free to send any number of points to the trustee within the range of 0 to 100. In the non-self-relevant condition, the participant took the role of one of two observers and watched another player send points to the trustee. The number of points that this other player sent was set randomly to be up to three points different from the number of points that the participant had sent in the practice game (in order to maximise similarity with participants' own allocations in the self-relevant condition). In both conditions, the trustee (Rabbit) returned only 1 point, thereby committing a trust violation.

Next, participants were told (for the first time) that one of the observers was now permitted either to reward or punish Rabbit by adding to or removing some of the accrued points. Participants then witnessed this observer remove either 53% (punishment condition) or 1% (no-punishment condition) of Rabbit's points. Participants then completed a questionnaire assessing their attitudes towards the punisher, Rabbit's deservingness, and their own schadenfreude and belief in a just world.

In Game 2, participants were assigned the role of observer (along with Rabbit), and saw a new player (Gorilla), transferring 50 points to the punisher from game 1, who acted as the trustee in this game and then returned 101 points. Participants were then given the option of punishing or rewarding this trustee (using the same procedure as in game 1). They could

choose a number between -100% to +100%, where -100% meant taking away all the trustee's points and +100% meant doubling the trustee's points. After making the decision, participants were debriefed.

Measures

Measures were mostly identical to those used in experiment 2. The four-item schadenfreude measure was again reliable (α = .87). The two items measuring participants' attitudes towards the third party were also highly correlated, $r(157) = .92$, $p < .001$.

Belief in a Just World (BJW). Participants rated two items from Lipkus'(1991) global BJW scale ("I feel that people get what they deserve" and "I basically feel that the world is a fair place") after Dolphin had delivered or not delivered punishment at the end of Game 1. Ratings on these items were positively correlated, $r(157) = .72$, $p < .001$, and were therefore combined into a composite score. Although BJW is most commonly assessed as a stable individual difference, this study focused on how it was affected by our Punishment and Self-Relevance manipulations.

Results

Schadenfreude, deservingness, and belief in a just world

Between-subjects 2 (Punishment: Punishment *vs*. No punishment) × 2 (Self-relevance: Self-relevant *vs*. Non-self-relevant) ANOVAs using schadenfreude, $F(1, 155) = 78.43$, $p < .001$, $\eta^2 = .336$, and deservingness, $F(1, 155) = 70.98$, $p < .001$, $\eta^2 = .314$, and BJW, $F(1, 155) = 7.57$, $p = .007$, $\eta^2 = .047$, as dependent measures all revealed significant main effects of punishment. Scores on all three measures were significantly higher in the punishment condition (schadenfreude $M = 3.93$, $SD = 1.44$; deservingness $M = 5.37$, $SD = 1.65$; BJW $M = 3.84$, $SD = 1.60$) than in the no punishment condition (schadenfreude $M = 2.10$, $SD = 1.16$; deservingness $M = 2.96$, $SD = 1.96$; BJW $M = 3.16$, $SD = 1.48$)[8]. There were

no other significant effects.

Attitudes

There was also a significant main effect of punishment on attitudes to the third party, $F(1, 155) = 86.38$, $p < .001$, $\eta^2 = .358$, which were more positive in the punishment condition ($M = 5.16$, $SD = 1.27$) than in the no punishment condition ($M = 3.10$, $SD = 1.52$). There were no other significant effects.

Behavior

A further significant main effect of punishment, $F(1, 155) = 21.94$, $p < .001$, $\eta^2 = .124$, showed that participants awarded the third party more points in the punishment condition ($M = 43.15$, $SD = 37.65$) than in the no punishment condition ($M = 12.47$, $SD = 45.87$). Self-relevance also had a significant main effect, $F(1, 155) = 10.86$, $p = .001$, $\eta^2 = .065$, showing that participants awarded more points in the non-self-relevant condition ($M = 38.88$, $SD = 40.43$) than in the self-relevant condition ($M = 17.01$, $SD = 45.98$), contrary to our prediction. The interaction effect was not significant.

Correlations

As shown in Table 2, schadenfreude was reliably positively correlated with perception of deservingness, BJW, and attitudes towards the third party in the punishment condition. However, as in experiment 2, neither schadenfreude nor attitudes was reliably correlated with the behavioral measure.

Serial mediation analysis

As self-relevance had no effect on the mediators and attitudes towards the third party, we used Hayes' (2022) PROCESS, model 6, to test the indirect effect of punishment on attitudes towards the third party via BJW, deservingness, and schadenfreude (see *Figure 2* for the path coefficients). Our results were consistent with the mediational model: punishment's total effect ($b = 2.07$, $SE = .22$, $t = 9.32$, $p < .001$) reduced with the direct effect remaining

significant (*b* = .80 , *SE* = .23, *t* = 3.48, *p* = .001) after controlling for the indirect effects (*indirect effect* = 1.27, *SE* = .21, 95% CI [.887 — 1.70]). The serial indirect effect through BJW, deservingness, and schadenfreude was significant (*indirect effect* = .02, *SE* = .01, 95% CI [.001 — .056]), consistent with our prediction. Table 3 shows all the indirect effects within the model.

Discussion

Experiment 3's findings were generally consistent with those of Experiments 1 and 2. First, participants again reported higher schadenfreude and deservingness in the punishment condition. Second, deservingness, schadenfreude, and attitudes towards the third-party were again positively intercorrelated in the punishment condition.

In Experiment 3, punishment was delivered more directly and did not involve a further violation of trust. As intended, this change resulted in participants having more favorable attitudes and behavior towards a third party who punished the trust violator than towards a third party who failed to punish the trust violator. This supports our previous conclusion that the less favorable reactions towards punishment in the previous experiments were due to their use of a further trust violation as a means of punishment.

Experiment 3 also included an additional measure of belief in a just world, and confirmed that participants' BJW was higher in the punishment condition. Our analyses also provided evidence consistent with our prediction that the effect of punishment on attitudes towards the third party were mediated serially via BJW, deservingness, and schadenfreude. This finding concurs with our theoretical account of schadenfreude as a corrective response to threats to participants' sense of justice presented by trust violation, and as a motivator encouraging alignment with the person upholding justice. Our results therefore suggest that the association between the trust violator's deservingness and schadenfreude may depend on

more general justice-related concerns.

Self-relevance significantly affected participants' reward behavior towards the third party but in the opposite direction to what we had predicted. Participants were in fact less generous towards the third- party punisher if their own trust rather than another player's trust had been violated in game 1. A likely explanation is that their personal experience of the original trust violation continued to diminish generosity to other players regardless of subsequent outcomes, consistent with previous research showing that victims of exploitation became less prosocial and more egoistic (Gollwitzer et al., 2009; Vohs et al., 2007).

Surprisingly, self-relevance had no other significant effects on our dependent measures. Participants did not report significantly greater schadenfreude, deservingness, or BJW, or more positive attitudes to a third-party punisher when their own rather than another player's trust was violated. Taken together, these findings suggest that participants' sense of justice was equally threatened by trust violations suffered by other players and those suffered by themselves.

Experiment 4

Experiment 3 provided novel evidence that schadenfreude's social function is related to the restoration of BJW by the third-party punisher. The study however had two important limitations. Firstly, the sampling plan and termination rules were solely based on Experiments 1 and 2, yielding a final sample size that was likely to be underpowered for the purposes of our complex serial mediation model. Secondly, we included other filler emotion measures and other attitudinal measures that were not intended for hypothesis testing. Following the requests of an anonymous reviewer of an earlier version of this paper, we therefore conducted a pre-registered replication of Experiment 3 using a larger sample as determined by a new power analysis for the serial mediation model.

Experiment 4 was identical to Experiment 3 except for the following modifications. Firstly, only the punishment manipulation was assessed as we did not find an effect of relevance in Experiment 3. Therefore, in both games, Dolphin was the third-party in Game 1 regardless of experimental condition. Secondly, we removed all tangential measures except for anger and gratitude towards the players in Game 1 and Game 2, which were retained to disguise the pre-programmed nature of the study and the particular focus on schadenfreude. Thirdly, with the removal of other measures, we were able to include all instead of only two items from the global BJW scale (Lipkus, 1991) to improve the reliability of the measurement[9]. Lastly, the visual aspects of the study were improved (e.g loading bars) to reduce any suspicions about the interactive nature of the study among participants. All hypotheses, data cleaning criteria, and analytical plans were pre-registered (https://osf.io/mzhbe/?view_only=c1113fa42c1948b290024dc8151575be) and were closely adhered to except that one more participant was recruited than intended due to technical delays.

Method

Participants

An a priori power analysis with Monte Carlo simulations (n = 1000) of parameter estimates obtained from Experiment 3 was conducted using the pwrSEM app (Wang & Rhemtulla, 2021) to determine the sample size required to conduct a serial mediation analysis with 80% power at an alpha level of .05. Results indicated that a sample size of 355 was needed to detect the expected serial indirect effects of punishment on attitudes. To account for possible failures in attention and suspicion checks, we pre-registered that 465 participants would be recruited.

Four hundred and sixty-six participants living in the UK were recruited through

Prolific (www.prolific.co) in exchange for £1.80 (~$2.25). Twenty-nine participants (6.22%) failed the attention check in Game 1 and sixty-one participants (13.09%) reported suspicions. They were excluded as specified in the pre-registered exclusion criteria, resulting in a final sample of 376 participants (213 women; 2 other, M_{age} = 36.63, *SD* = 12.79).

Procedure and Materials

As in Experiment 3, participants were first familiarised with the experimental task. They were randomly exposed to the punishment (*n* = 186) or non-punishment (*n* = 190) manipulation in Game 1. Participants then completed a questionnaire (see Experiment 3) consisting of attention and manipulation checks ("How strongly did Dolphin reward or punish Rabbit*"*, 1 = "*A very big punishment*" to 11 = "*A very big reward*"), four filler emotion items (anger and gratitude towards Rabbit (trust violator) and Dolphin (third-party)), and measures of BJW (α = .89), deservingness, and schadenfreude (α = .90), and attitudes towards Dolphin, $r(374) = .93$, $p < .001$. Lastly, they took part in Game 2 and decided the extent to which they would punish or reward Dolphin. They also completed a questionnaire about Game 2 consisting of similar filler emotion items and attitudinal measures about Gorilla and Dolphin that were not intended for any analysis. The inclusion of these items was to make Game 2 appear similar to Game 1 and thereby disguise our particular interest in events happened in Game 1.

Results

Manipulation check.

Dolphin was considered to have delivered a stronger punishment in the punishment condition (*M* = 4.01, *SD* = 1.78) than in the non-punishment condition (*M* = 5.75, *SD* = 1.80), $t(374) = 9.43$, $p < .001$, $d = .97$, indicating that the punishment manipulation was effective.

Main analysis

Following the pre-registered analytical plan, a one-way MANOVA revealed an overall significant difference between punishment and no punishment conditions on the key dependent variables (BJW, deservingness, schadenfreude, attitudes towards Dolphin, and returned points), $V = .53$, $F(5, 370) = 83.23$, $p < .001$, $\eta^2 = 0.53$. Follow-up univariate tests with Bonferroni correction revealed that compared with no punishment, punishment resulted in significantly higher BJW ($M = 3.72$, $SD = 1.16$ vs. $M = 2.89$, $SD = 1.13$), $F(374) = 50.49$, $p < .001$, $\eta^2 = 0.12$; deservingness ($M = 5.45$, $SD = 1.39$ vs. $M = 3.30$, $SD = 2.06$), $F(374) = 141.99$, $p < .001$, $\eta^2 = .28$; schadenfreude ($M = 4.36$, $SD = 1.29$ vs. $M = 1.99$, $SD = 1.23$); $F(374) = 332.69$, $p < .001$, $\eta^2 = .47$; attitudes towards third-party ($M = 5.21$, $SD = 1.02$, $M = 2.89$, $SD = 1.40$), $F(374) = 337.34$, $p < .001$, $\eta^2 = .47$; and significantly more points returned to the third party($M = 27.37$, $SD = 34.98$ vs. $M = 17.03$, $SD = 40.92$), $F(374) = 6.94$, $p = .044$, $\eta^2 = .02$. Table 2 shows the correlations between all measures.

Serial Mediation Analysis

Using Hayes' (2022) PROCESS (Model 6) with 10,000 random bootstrap samples at 95% CI, it was shown that the total effect of punishment on attitudes ($b = 2.32$, $p < .001$) was reduced ($b = 0.73$, $p < .001$) after controlling for the indirect effects (*indirect effect* = 1.59, *SE* = .12, 95% CI [1.35 — 1.86]). Along with other indirect effects (See Table 3), the serial indirect effect through BJW, deservingness, and schadenfreude was significant (*indirect effect* = .02, *SE* = .01, 95% CI [.007 — .044]), replicating Experiment 3 and consistent with the proposed theoretical model.

Exploratory Analysis

Although it was not a primary concern during the preregistration stage, in response to an anonymous reviewer’s concern that positive emotional responses other than schadenfreude might better explain the positive social effects of punishment on attitudes, we utilised this better-powered study to explore gratitude towards the third-party after Game 1 as an

alternative explanation for the present findings. Gratitude is a positive emotion directed to the benefactor who has done something intentionally that benefits the recipient, and its social affiliative function is well-documented (Bartlett et al., 2012; McCullough et al., 2008). Therefore, it seemed possible that participants might have a more positive attitudes towards the third-party because they felt grateful to them for the punishment that they had delivered. However, an independent t-test revealed no significant difference in rated gratitude between punishment and non-punishment condition (p = .149). Further, serial mediation analysis with gratitude as a covariate did not affect the serial indirect effects of BJW, deservingness, and schadenfreude in the relationship between punishment and attitudes (*indirect effect* = .01, *SE* = .01, 95% CI [.002 — .022]), suggesting that gratitude could not fully explain the findings in the main analyses.

Unlike in Experiments 1 to 3, punishment manipulation had a significant effect on points returned and there were small but significant relationships amongst schadenfreude, attitude, and points returned. These led to the possibility that the positive effect of punishment on attitudes to the third party might be translated into reward. Although this chain of effects was not pre-registered, we conducted exploratory analysis to assess whether results were consistent with a model where the effect of punishment on points returned to the third-party was serially mediated through BJW, deservingness, schadenfreude and attitude. Using Hayes' (2022) PROCESS (Model 6; 10000 bootstrap samples), we found that the total effect of punishment on behavior (b = 10.34, p = .009) became non-significant (b = 0.12, p = .983), after controlling for indirect effects (*indirect effect* = 10.22, *SE* = 4.35, 95% CI [1.856 — 18.865]). Further, the indirect effect with the four serial mediators was also significant (*indirect effect* = 0.17, *SE* = .010, 95% CI [.033 — .417]), consistent with a full mediation model (See Table 4). See *Figure* 3 for all path coefficients.

Discussion

Replicating Experiment 3, participants in the punishment condition reported higher levels of BJW, deservingness, and schadenfreude and had more positive attitudes and returned more points to the third-party than in the no punishment condition. The results of the mediation analysis were again consistent with a model where punishment's effect on attitudes towards the third-party involved the three proposed mediators. Though exploratory in nature, we also extended Experiment 3 by obtaining results consistent with a model where the effect of punishment (vs. no punishment) on points returned was fully and serially mediated by BJW, deservingness, schadenfreude, and attitudes towards the third-party. Controlling for the experience of gratitude did not change the conclusions from these findings. Taken together, this high-powered, pre-registered experiment provided robust support for the proposed social affiliative function of schadenfreude.

General Discussion

The findings of these four experiments extend our understanding of schadenfreude in relation to perceived justice and of the social consequences of this emotion. Our punishment manipulation had consistent effects on schadenfreude and deservingness across experiments. As expected, participants reported stronger schadenfreude and greater perceptions of deservingness after observing a player who had previously violated their trust received a negative outcome from the third party. Manipulating trust violations in economic games was apparently a reliable method for inducing feelings of schadenfreude. Similar methodologies may help to circumvent some of the limitations of vignette-based designs in schadenfreude research (Parkinson & Manstead, 1993, 2015).

Positive correlations between deservingness and schadenfreude in the punishment condition were also consistent across experiments, confirming the results of previous research

(e.g., Feather & Sherman, 2002). Further, both schadenfreude and deservingness were consistent positive predictors of participants' attitudes towards the third-party punisher. In addition, the results of Experiment 1 were consistent with past research showing that deservingness mediated the effect of punishment on schadenfreude, (Feather & Mckee, 2014; Feather & Sherman, 2002).

The designs of our experiments also helped clarify the boundary conditions for the effect of punishment on schadenfreude and interpersonal attitudes towards third parties. Experiment 1 showed that deserved punishment led to stronger schadenfreude, but not to improved attitudes towards third parties because of the associated trust violation. Experiment 2 showed that deserved punishment led to stronger schadenfreude and relatively more favorable attitudes towards a third party who delivered a negative outcome to the trust violator rather than another player. Further, the results of our mediation analysis were consistent with a model where this effect of punishment on attitudes was serially mediated by deservingness and schadenfreude after removing the negative effect of the trust violation associated with delivering a punishment, suggesting that schadenfreude and its corresponding interpersonal effects are restricted to the delivery of retributive justice rather than any negative outcomes befalling an unrelated person.

By separating punishment from trust violation, Experiment 3 further showed that punishment directly influenced both schadenfreude and attitudes towards third parties regardless of whether the prior trust violation was committed against the participant or another player. Experiment 3 also produced results that were consistent with our serial mediation model in which punishment predicted higher levels of just-world beliefs, deservingness, schadenfreude, and more positive attitudes towards the punisher. Further, evidence consistent with this chain of indirect effects was replicated in a high-powered, pre-registered Experiment 4 and exploratory analysis suggested that this serial mediation might

also apply to positive behavior towards the third-party punisher. Hence, our research extends previous findings relating schadenfreude to just world beliefs (Pietraszkiewicz, 2013) by directly demonstrating that the restorative effect of punishment on these beliefs can lead to schadenfreude, regardless of whether the trust violation happened to oneself (Experiments 3 & 4) or another person (Experiment 3).

By including BJW, the findings of Experiments 3 and 4 extend our theoretical understanding of how the perceived deservingness of a misfortune might result in the experience of schadenfreude. In particular, a small body of research shows that people at times feel schadenfreude about a target experiencing an objectively undeserved misfortune if they reconstrue the situation from "undeserved" to "deserved" (e.g. Berndsen et al., 2017). Drawing insights from the BJW literature (cf. Ellard et al., 2016), perhaps this change in deservingness judgment could be understood as a coping strategy to maintain the perception of justice, keeping one's BJW at a homeostatic level. This view of BJW being a homeostatic system fits well with the proposition that schadenfreude is a response to the restoration of an imbalanced situation (van de Ven, 2018). Specifically, if a justice-related misfortune results in a desirably high level of BJW, the observer will evaluate the event as deserved and consequently experience schadenfreude. Correspondingly, if a justice-related misfortune results in an undesirably low level of BJW, the observer will evaluate the event as undeserved and consequently experience sympathy (Feather & Sherman, 2002).

Theoretical interpretations aside, there are potential questions about whether the three key mediators, BJW, deservingness, and schadenfreude, are empirically distinguishable, especially in the punishment condition where the serial mediation model that we used might seem to imply a clear separability of these variables. However, zero-order correlations in Experiments 3 and 4 (See Table 2) showed that the intercorrelations between these constructs were at most moderate (Cohen, 1988) and did not exceed various suggested cutoffs of

correlation at .75 (Voorhees et al., 2016) or .80 (Rönkkö & Cho, 2022), indicating there were no serious issues of discriminant validity.[10] More importantly, controlling for deservingness (see footnotes 3, 6, and 8) did not remove the significant effect of punishment on schadenfreude, confirming statistical independence of the obtained differences in scores on these variables across experimental conditions. Future research should collect additional data concerning the distinctions between the key variables assessed in this study and systematically evaluate alternative models that may underlie the relationships between BJW, deservingness, and schadenfreude (see also below).

Another theoretical contribution of the present research relates to the growing interest in the social functions of schadenfreude (e.g. Lange & Boecker, 2019). Unlike other studies in which BJW or deservingness were examined as elicitors of schadenfreude, we are the first researchers – to the best of our knowledge – to show that justice-based schadenfreude serves an affiliative function that helps to align participants with a third-party punisher, creating an alliance that defends justice against the threat posed by a trust violator. This offers individuals an efficient way to establish interpersonal relationships with those who share the same belief in justice and to bolster their confidence that others will uphold the associated social norms. Although schadenfreude is experienced and judged by others as a negative emotion, the present results suggest that it can serve positive functions in moral life.

However, we do not wish to claim this function of schadenfreude generalizes across all possible contexts. We concur with van Dijk et al.'s (2015) and Wang et al.'s (2019) conclusion that schadenfreude is a multifaceted, multipurpose emotion. In some situations, schadenfreude may serve a socially distancing rather than affiliative function. For example, schadenfreude experienced after a contestant on TV receives negative feedback (van Dijk et al., 2012) or when a rival soccer team loses an important match (Leach & Spears, 2009) may serve to relieve the pain of inferiority by creating or enhancing status differentials between

people or groups (Fischer & Manstead, 2016; van Dijk et al., 2015). Unlike the affiliative and solidarity functions emphasised in the present research, these outcomes relate more directly to the common perception of the "darker side" of schadenfreude (Cikara et al., 2011; Schumpe & Lafrenière, 2016), experienced by individuals with higher levels of the Dark Triad of personality (James et al., 2014) or in more competitive settings (Leach & Spears, 2009).

Finally, we found through exploratory analyses that punishment did not affect gratitude towards the third-party and that the findings that were consistent with the hypothesised sequential effects of BJW, deservingness, and schadenfreude remained the same even after controlling for gratitude. This might mean that the effect of punishment on attitudes was due to seeing retributive justice being served to the trust-violator by the third-party (i.e. schadenfreude) rather than a result of directly benefitting from the third-party's intervention (i.e. gratitude). Our research therefore contributes to the ongoing debate about whether schadenfreude is a unique emotion that can be distinguished from other positive emotions (Hess, 2018). In our view, the kinds of schadenfreude and gratitude experienced in the context investigated here may differ in the relative focus of their social targets (second-party vs third-party) as well as the concerns for upholding norms and justice (more vs. less). This interpretation aligns with recent research showing that gratitude draws attention to relationship building at the expense of concerns for justice (Zhu et al., 2020). Nevertheless, it is important to stress that we do not think schadenfreude was the only emotional response to deserved punishment experienced by participants in these experiments or the only explanation for why they aligned with the third-party. Other emotions such as contentment (Cordaro et al., 2016) or elevation (Siegel et al., 2014) may co-occur and foster the social bond between the self and the third-party. Future research should attempt to disentangle the specific effects of schadenfreude from the effects of other positive emotions.

Limitations

The present findings should be interpreted with several limitations in mind. First, our mediational analyses were based on variables measured cross sectionally rather than longitudinally, leaving open the possibility that alternative causal pathways might explain the current findings. For instance, the relationships between deservingness and emotion (Feather, 2006) and between BJW and schadenfreude (Wang et al., 2019) could operate bidirectionally. Hence, caution should be taken in drawing any firm causal conclusions from the results of our serial mediation analysis because alternative models might also result in acceptable levels of statistical fit with the present data. Future research should follow the manipulation-of-mediator method advocated by Spencer et al. (2005) and directly manipulate the mediators with random assignment to establish a causal pathway between punishment, the serial mediators, and attitudes more conclusively.

Second, although our experiments used an involving real-time task rather than imagined scenarios or recalled experiences, our participants interacted remotely rather than directly during the trust games investigated here. This allowed us to control the series of outcomes to which participants responded, but also brought the disadvantage of reducing the level of interpersonal engagement during the task in Experiments 2 – 4. However, the consistency of findings obtained in Experiment 1 where participants actually met co-players and Experiments 2-4 where all interaction operated virtually suggests that the level of contact between interactants does not make a clear difference to the processes we were investigating in this research. Nonetheless, we acknowledge that the experimental contexts used in the present research did not match all aspects of the most common episodes of schadenfreude experienced outside the laboratory. One possible improvement would be to have participants play games face-to-face in real time, thus permitting more direct assessment of moment-by-moment causes and consequences of the experience and expression of schadenfreude.

However, this would also mean a reduction in the level of experimental control that was possible.

A final limitation is that we relied on self-report measures to assess many key constructs, raising concerns about whether participants gave accurate accounts of operative mental states and processes. Alleviating some of these concerns, we included participants' behavior in the final game of all experiments to confirm that punishment did not only affect self-reports. However, with the exception of Experiment 4, there were no consistent significant associations between this behavioral measure and self-reported attitudes, deservingness or schadenfreude in the punishment conditions. A possible explanation is that behavior was influenced not only by attitudes towards the trustee, but also by other factors such as social value orientation, risk attitudes, and fairness norms, which have been found to be influential in complex experimental designs such as those used in the present research (Bicchieri & Mercier, 2013; Kanagaretnam et al., 2009). Further, some of our participants might have seen the final game as an opportunity to model appropriate trust-game behavior to other players (including trust violators and punishers) meaning that their behavior did not only depend on their feelings towards the current investor. Future research should investigate the effects of these possible extraneous factors and compare trust-game outcomes with other behavioral measures including those derived from other experimental games (e.g., dictator games and Prisoners' dilemma).

Conclusions

These experiments demonstrate that schadenfreude can be induced in interpersonal trust games and clarify some of this emotion's consequences for interpersonal relations in a justice-related context. We showed that schadenfreude is associated with improved attitudes towards people inflicting deserved punishment, but that this effect may be counteracted when

punishment also involves trust violation. In contexts where punishment is more explicitly retributive, attitude improvement apparently depends on the desire to restore belief in a just-world.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423. https://doi.org/10.1037/0033-2909.103.3.411

Bargh, J., & Chartrand, T. (2000). A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253–285.). Cambridge University Press.

Bartlett, M. Y., Condon, P., Cruz, J., Baumann, J., & Desteno, D. (2012). Gratitude: Prompting behaviours that build relationships. *Cognition and Emotion*, *26*(1). https://doi.org/10.1080/02699931.2011.561297

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Berndsen, M., & Feather, N. T. (2016). Reflecting on schadenfreude: Serious consequences of a misfortune for which one is not responsible diminish previously expressed schadenfreude; the role of immorality appraisals and moral emotions. *Motivation and Emotion*, *40*(6), 895–913. https://doi.org/10.1007/s11031-016-9580-8

Berndsen, M., Tiggemann, M., & Chapman, S. (2017). "It wasn't your fault, but …...": Schadenfreude about an undeserved misfortune. *Motivation and Emotion*, *41*(6), 741–748. https://doi.org/10.1007/s11031-017-9639-1

Bicchieri, C., & Mercier, H. (2013). Self-serving biases and public justifications in trust games. *Synthese*, *190*(5), 909–922. https://doi.org/10.1007/s11229-012-0192-x

Bies, R. J., & Tripp, T. M. (1996). Beyond distrust: "Getting even" and the need for revenge. In *Trust in Organizations: Frontiers of Theory and Research* (pp. 246–260). SAGE Publications, Inc. https://doi.org/10.4135/9781452243610.n12

Brambilla, M., & Riva, P. (2017). Self-image and schadenfreude: Pleasure at others'

misfortune enhances satisfaction of basic human needs. *European Journal of Social Psychology*, *47*(4), 399–411. https://doi.org/10.1002/ejsp.2229

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, *83*(2), 284–299. https://doi.org/10.1037/0022-3514.83.2.284

Carlsmith, K. M., Wilson, T. D., & Gilbert, D. T. (2008). The paradoxical consequences of revenge. *Journal of Personality and Social Psychology*, *95*(6), 1316–1324. https://doi.org/10.1037/a0012165

Charness, G., Cobo-Reyes, R., & Jiménez, N. (2008). An investment game with third-party intervention. *Journal of Economic Behavior and Organization*, *68*(1), 18–28. https://doi.org/10.1016/j.jebo.2008.02.006

Cikara, M., Botvinick, M. M., & Fiske, S. T. (2011). Us versus them: Social identity shapes neural responses to intergroup competition and harm. *Psychological Science*, *22*(3), 306–313. https://doi.org/10.1177/0956797610397667

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Academic Press.

Cordaro, D. T., Brackett, M., Glass, L., & Anderson, C. L. (2016). Contentment: Perceived completeness across cultures and traditions. *Review of General Psychology*, *20*(3). https://doi.org/10.1037/gpr0000082

Crockett, M. J., Özdemir, Y., & Fehr, E. (2014). The value of vengeance and the demand for deterrence. *Journal of Experimental Psychology: General*, *143*(6), 2279–2286. https://doi.org/10.1037/xge0000018

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, *7*(4), 324–336. https://doi.org/10.1207/S15327957PSPR0704_05

Ellard, J. H., Harvey, A., & Callan, M. J. (2016). The Justice Motive: History, Theory, and Research. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of Social Justice Theory and Research* (pp. 127–143). Springer New York. https://doi.org/10.1007/978-1-4939-3216-0_7

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*(2), 175–191. https://doi.org/10.3758/BF03193146

Feather, N. T. (2006). Deservingness and emotions: Applying the structural model of deservingness to the analysis of affective reactions to outcomes. *European Review of Social Psychology*, *17*(1), 38–73. https://doi.org/10.1080/10463280600662321

Feather, N. T. (2014). Deservingness and schadenfreude. In W. W. van Dijk & J. W. Ouwerkerk (Eds.), *Schadenfreude: Understanding Pleasure at the Misfortune of Others* (pp. 29–57). https://doi.org/10.1007/9781139084246.004

Feather, N. T., & Mckee, I. R. (2014). Deservingness, liking relations, schadenfreude, and other discrete emotions in the context of the outcomes of plagiarism. *Australian Journal of Psychology*, *66*(1), 18–27. https://doi.org/10.1111/ajpy.12030

Feather, N. T., & Sherman, R. (2002). Envy, resentment, schadenfreude, and sympathy: Reactions to deserved and undeserved achievement and subsequent failure. *Personality and Social Psychology Bulletin*, *28*(7), 953–961. https://doi.org/10.1177/014616720202800708

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140. https://doi.org/10.1038/415137a

Fischer, A. H., & Manstead, A. S. R. (2016). Social functions of emotions. In L. F. Barrett,

M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of Emotions* (4th ed., pp. 424–439). Guilford.

Gollwitzer, M., Rothmund, T., Pfeiffer, A., & Ensenbach, C. (2009). Why and when justice sensitivity leads to pro- and antisocial behavior. *Journal of Research in Personality, 43*(6), 999–1005. https://doi.org/10.1016/j.jrp.2009.07.003

Gordon, D. S., Madden, J. R., & 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*(10), e110045. https://doi.org/10.1371/journal.pone.0110045

Hafer, C. L. (2000). Do innocent victims threaten the belief in a just world? Evidence from a modified Stroop task. *Journal of Personality and Social Psychology, 79*(2), 165–173. https://doi.org/10.1037/0022-3514.79.2.165

Hafer, C. L., Drolet, C. E., Davis, E. E., Segalowitz, S. J., & Shulman, E. P. (2020). Evidence of a processing advantage for deservingness-relevant information. *Social Psychology*, *51*(2), 127–134. https://doi.org/10.1027/1864-9335/a000396

Hafer, C. L., & Rubel, A. N. (2015). The why and how of defending belief in a just world. *Advances in Experimental Social Psychology*, *51*(1), 41–96. https://doi.org/10.1016/bs.aesp.2014.09.001

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (3rd Edition)*. Guildford.

Hess, U. (2018). Why are schadenfreude and gluckschmerz not happiness or anger? Or are they? *Emotion Review*, *10*(4), 306–308. https://doi.org/10.1177/1754073918790027

James, S., Kavanagh, P. S., Jonason, P. K., Chonody, J. M., & Scrutton, H. E. (2014). The Dark Triad, schadenfreude, and sensational interests: Dark personalities, dark emotions, and dark behaviors. *Personality and Individual Differences*, *68*.

https://doi.org/10.1016/j.paid.2014.04.020

Kanagaretnam, K., Mestelman, S., Nainar, K., & Shehata, M. (2009). The impact of social value orientation and risk attitudes on trust and reciprocity. *Journal of Economic Psychology*, *30*(3), 368–380. https://doi.org/10.1016/j.joep.2008.12.003

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, *13*(5), 505–521. https://doi.org/10.1080/026999399379168

Kiyonari, T., & Barclay, P. (2008). Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology*, *95*(4), 826–842. https://doi.org/10.1037/a0011381

Lange, J., & Boecker, L. (2019). Schadenfreude as social-functional dominance regulator. *Emotion*, *19*(3), 489–502. https://doi.org/10.1037/emo0000454

Leach, C. W., & Spears, R. (2009). Dejection at in-group defeat and schadenfreude towards second- and third-party out-groups. *Emotion*, *9*(5), 659–665. https://doi.org/10.1037/a0016815

Lerner, M. J. (1980). *The belief in a just world*. Springer US. https://doi.org/10.1007/978-1-4899-0448-5

Lipkus, I. (1991). The construction and preliminary validation of a global belief in a just world scale and the exploratory analysis of the multidimensional belief in a just world scale. *Personality and Individual Differences*, *12*(11), 1171–1178. https://doi.org/10.1016/0191-8869(91)90081-L

McCullough, M. E., Kimeldorf, M. B., & Cohen, A. D. (2008). An adaptation for altruism. *Current Directions in Psychological Science*, *17*(4). https://doi.org/10.1111/j.1467-8721.2008.00590.x

Nelissen, R. M. A. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior*, *29*(4), 242–248.

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

Olson, J. M., & Hafer, C. L. (2001). Tolerance of Personal Deprivation. In J. T. Jost & B. Major (Eds.), *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations* (pp. 157–175). Cambridge University Press.

Parkinson, B., Fischer, A. H., & Manstead, A. S. R. (2005). Emotion in social relations: Cultural, group, and interpersonal processes. In *Emotion in Social Relations: Cultural, Group, and Interpersonal Processes*. https://doi.org/10.4324/9780203644966

Parkinson, B., & Manstead, A. S. R. (1993). Making sense of emotion in stories and social life. *Cognition and Emotion*, *7*(3–4), 295–323. https://doi.org/10.1080/02699939308409191

Parkinson, B., & Manstead, A. S. R. (2015). Current emotion research in social psychology: Thinking about emotions and other people. *Emotion Review*, *7*(4), 371–380. https://doi.org/10.1177/1754073915590624

Pietraszkiewicz, A. (2013). Schadenfreude and just world belief. *Australian Journal of Psychology*, *65*(3), 188–194. https://doi.org/10.1111/ajpy.12020

Portmann, J. (2002). *When bad things happen to other people*. Routledge.

Raihani, N. J., & Bshary, R. (2015). The reputation of punishers. *Trends in Ecology and Evolution*, *30*(2), 98–103. https://doi.org/10.1016/j.tree.2014.12.003

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404. https://doi.org/10.5465/AMR.1998.926617

Schindler, R., Körner, A., Bauer, S., Hadji, S., & Rudolph, U. (2015). Causes and

consequences of schadenfreude and sympathy: A developmental analysis. *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0137669

Schumpe, B. M., & Lafrenière, M. A. K. (2016). Malicious joy: Sadism moderates the relationship between schadenfreude and the severity of others' misfortune. *Personality and Individual Differences*, *94*. https://doi.org/10.1016/j.paid.2016.01.005

Siegel, J. T., Thomson, A. L., & Navarro, M. A. (2014). Experimentally distinguishing elevation from gratitude: Oh, the morality. *Journal of Positive Psychology*, *9*(5). https://doi.org/10.1080/17439760.2014.910825

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*(6). https://doi.org/10.1037/0022-3514.89.6.845

Sundie, J. M., Ward, J. C., Beal, D. J., Chin, W. W., & Geiger-Oneto, S. (2009). Schadenfreude as a consumption-related emotion: Feeling happiness about the downfall of another's product. *Journal of Consumer Psychology*, *19*(3). https://doi.org/10.1016/j.jcps.2009.02.015

Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, *77*, 184–197. https://doi.org/10.1016/j.chb.2017.08.038

van de Ven, N. (2018). Schadenfreude and gluckschmerz are emotional signals of (im)balance. *Emotion Review*, *10*(4), 305–306. https://doi.org/10.1177/1754073918768883

van Dijk, W. W., Goslinga, S., & Ouwerkerk, J. W. (2008). Impact of responsibility for a misfortune on schadenfreude and sympathy: Further evidence. *Journal of Social Psychology*, *148*(5), 631–636. https://doi.org/10.3200/SOCP.148.5.631-636

van Dijk, W. W., & Ouwerkerk, J. W. (2014). Schadenfreude: Understanding pleasure at the misfortune of others. In W. W. van Dijk & J. W. Ouwerkerk (Eds.), *Schadenfreude: Understanding Pleasure at the Misfortune of Others*. Cambridge University Press. https://doi.org/10.1017/CBO9781139084246

van Dijk, W. W., Ouwerkerk, J. W., Smith, R. H., & Cikara, M. (2015). The role of self-evaluation and envy in schadenfreude. *European Review of Social Psychology*, *26*(1), 247–282. https://doi.org/10.1080/10463283.2015.1111600

van Dijk, W. W., Ouwerkerk, J. W., van Koningsbruggen, G. M., & Wesseling, Y. M. (2012). "So you wanna be a pop star?": Schadenfreude following another's misfortune on TV. *Basic and Applied Social Psychology*, *34*(2). https://doi.org/10.1080/01973533.2012.656006

Vohs, K. D., Baumeister, R. F., & Chin, J. (2007). Feeling duped: Emotional, motivational, and cognitive Aspects of being exploited by others. *Review of General Psychology*, *11*(2), 127–141. https://doi.org/10.1037/1089-2680.11.2.127

Voorhees, C. M., Brady, M. K., Calantone, R., & Ramirez, E. (2016). Discriminant validity testing in marketing: an analysis, causes for concern, and proposed remedies. *Journal of the Academy of Marketing Science*, *44*(1), 119–134. https://doi.org/10.1007/s11747-015-0455-4

Wang, S., Lilienfeld, S. O., & Rochat, P. (2019). Schadenfreude deconstructed and reconstructed: A tripartite motivational model. *New Ideas in Psychology*, *52*, 1–11. https://doi.org/10.1016/j.newideapsych.2018.09.002

Wang, Y. A., & Rhemtulla, M. (2021). Power analysis for parameter estimation in structural equation modeling: A discussion and tutorial. *Advances in Methods and Practices in Psychological Science*, *4*(1). https://doi.org/10.1177/2515245920918253

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the Web: How unattended

selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, *111*(4), 493–504. https://doi.org/10.1037/pspa0000056

Zhu, R., Xu, Z., Tang, H., Wang, H., Zhang, S., Zhang, Z., Mai, X., & Liu, C. (2020). The dark side of gratitude: Gratitude could lead to moral violation. *Journal of Experimental Social Psychology*, *91*. https://doi.org/10.1016/j.jesp.2020.104048

Table 1

Structure of games in Experiment 2

GAME 1: Trust violation	
Players	**Outcome**
Investor: Participant *Trustee*: Trust Violator	Trustee fails to return points requested by investor
GAME 2a: Punishment manipulation	
Players	**Outcome**
Investor: Trust Violator *Trustee*: Third Party	Trustee either returns or fails to return points requested by investor (depending on punishment condition)
GAME 2b: Matched outcomes	
Players	**Outcome**
Investor: New Player 1 *Trustee*: New Player 2	Trustee either returns or fails to return points requested by investor (depending on punishment condition)
GAME 3: Final investor manipulation (and behavioral measure)	
Players	**Outcome**
Investor: Third Party or New Player 2 (Different Investor) *Trustee*: Participant	Number of points returned by participant provides behavioral measure

Table 2

Correlations in Experiments 3 and 4's Punishment Condition

	Experiment	Deservingness	Schadenfreude	Attitudes	Behavior
Belief in a Just World	Exp 3	.27*	.26*	.26*	-.04
		[.053 .457]	[.047, .453]	[.046, .452]	[-.252, .182]
	Exp 4	.16*	.29**	.25**	-.13
		[.023, .300]	[.153, .414]	[.116, .383]	[-.270, .010]
Deservingness	Exp 3	-	.43**	.28*	.10
			[.238, .594]	[.067, .468]	[-.124, .306]
	Exp 4		.41**	.38**	.15*
			[.286 .523]	[.252, .496]	[.009, .287]
Schadenfreude	Exp 3		-	.52**	.01
				[.334, .657]	[-.212, .222]
	Exp 4			.61**	.03
				[.513, .693]	[-.114, .171]
Attitudes	Exp 3			-	-.04
					[-.258 .175]
	Exp 4				.06
					[-.087, .197]

Note. Experiment 3: $N = 82$; Experiment 4: $N = 190$. Numbers in brackets represent 95% confidence intervals. * Significant at the $p < .05$ level (2-tailed); ** Significant at the $p < .01$ level (2-tailed).

Table 3
Indirect effect of condition on attitudes towards third-party

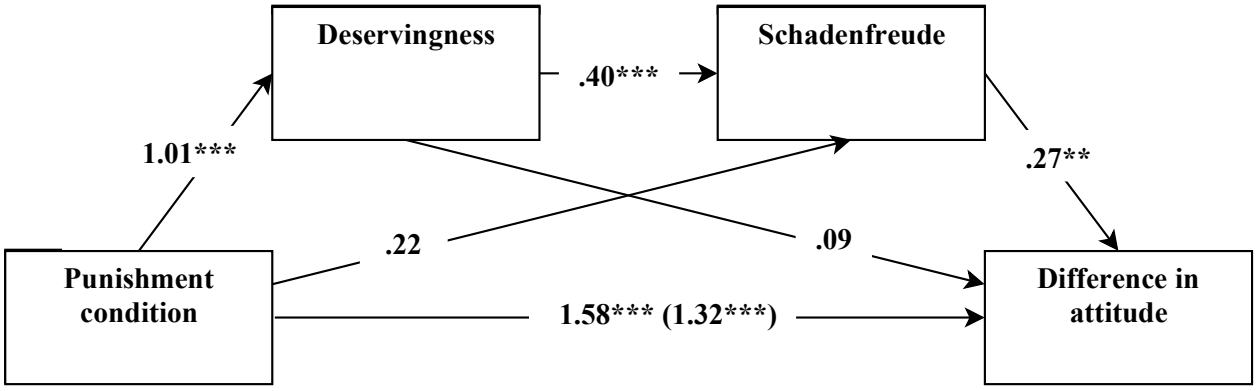
Indirect Effects	Experiment 3			Experiment 4		
	Effect	*SE*	95%CI	Effect	*SE*	95% CI
Condition → BJW → Attitudes	0.09	.06	[-.002, .024]	0.07	0.02	**[.023, .121]**
Condition → Deservingness → Attitudes	0.29	0.16	**[.007, .663]**	0.21	0.05	**[.117, .315]**
Condition → Schadenfreude → Attitudes	0.46	0.16	**[.179, .792]**	0.41	0.06	**[.296, .536]**
Condition → BJW→ Deservingness →Attitudes	0.02	0.02	[-.001, .065]	0.02	0.01	**[.006, .039]**
Condition → BJW→ Schadenfreude →Attitudes	0.05	0.03	**[.003, .119]**	0.08	0.02	**[.045, .114]**
Condition → Deservingness →Schadenfreude →Attitudes	0.34	0.10	**[.168, .553]**	0.14	0.03	**[.092, .192]**
Condition → BJW→Deservingness → Schadenfreude → Attitudes	0.02	0.01	**[.001, .056]**	0.01	0.01	**[.004, .025]**

BJW = Belief in a Just World; *N* (Experiment 3) = 159; *N* (Experiment 4) = 376;

Table 4
Indirect effect of condition on points returned to third-party in Experiment 4

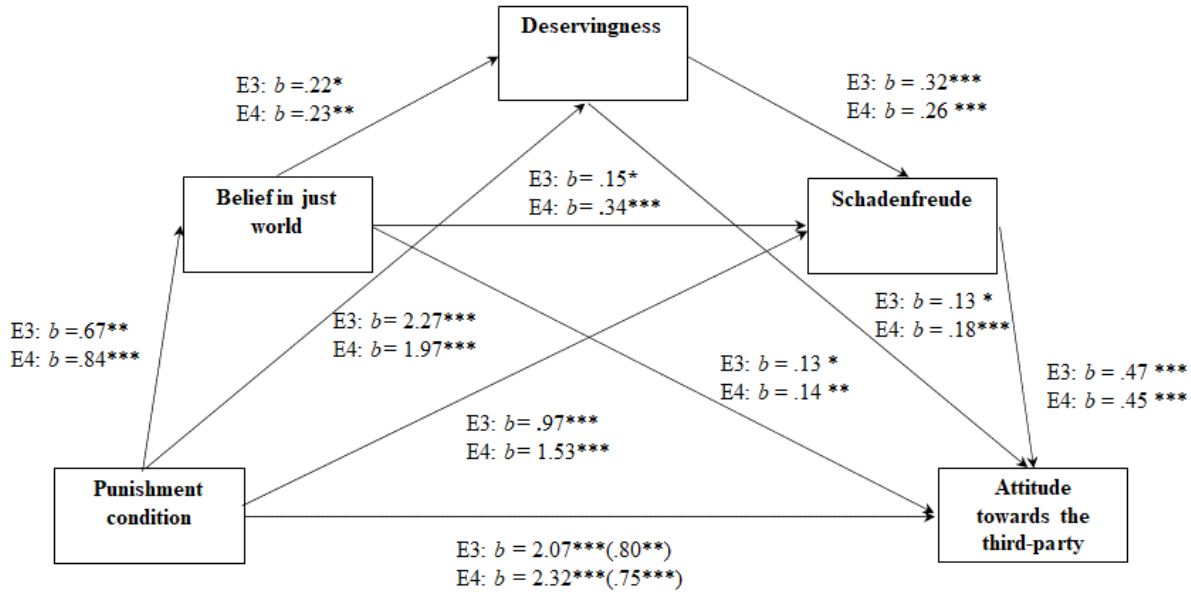
	Effect	*SE*	95% CI
Condition → BJW → Behavior	-2.34	1.65	[-5.825, .571]
Condition → Deservingness → Behavior	-1.21	2.51	[-6.425, 3.469]
Condition → Schadenfreude → Behavior	-2.58	3.25	[-9.077, 3.723]
Condition → Attitudes → Behavior	5.76	2.04	**[2.248, 10.195]**
Condition → BJW→ Deservingness → Behavior	-0.12	0.26	[-.675, .360]
Condition → BJW→ Schadenfreude → Behavior	-0.48	0.62	[-1.759, .688]
Condition → BJW→ Attitudes → Behavior	0.90	0.47	**[.193, 1.998]**
Condition → Deservingness → Schadenfreude → Behavior	-0.86	1.11	[-3.185, 1.221]
Condition → Deservingness → Attitudes → Behavior	2.68	1.14	**[.920, 5.347]**
Condition → Schadenfreude →Attitudes → Behavior	5.35	1.78	**[2.156, 9.146]**
Condition → BJW→ Deservingness → Schadenfreude → Behavior	-0.08	0.12	[-.369, .117]
Condition → BJW→ Deservingness → Attitudes → Behavior	0.26	0.16	**[.048, .649]**
Condition → BJW→ Schadenfreude → Attitudes → Behavior	0.99	0.40	**[.348, 1.897]**
Condition → Deservingness → Schadenfreude → Attitudes → Behavior	1.79	0.67	**[.674, 3.259]**
Condition → BJW → Deservingness → Schadenfreude → Attitudes → Behavior	0.17	0.10	**[.033, .417]**

Figure 1. Indirect effect of punishment on difference in attitudes towards the third-party and towards the different investor via deservingness and schadenfreude in Experiment 2



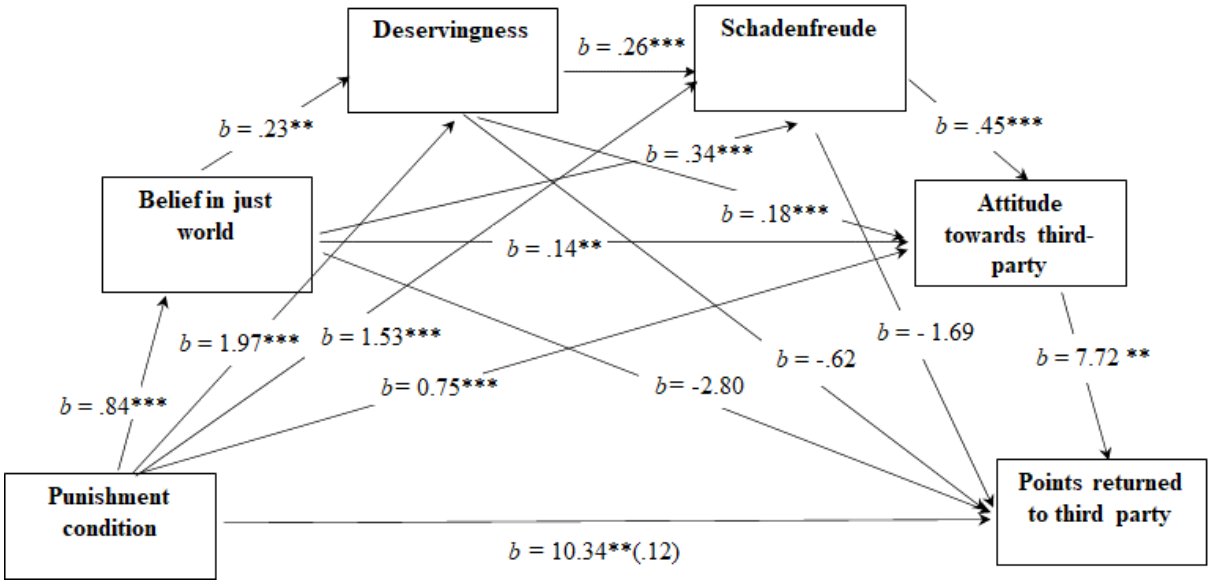
The punishment conditions were dummy coded (0: No punishment; 1: Punishment) All values represent unstandardised coefficients. *$p < .05$; **$p < .01$; ***$p < .001$

Figure 2. Indirect effects of punishment on participants' attitudes towards the third-party punisher via belief in the just world, deservingness and schadenfreude in Experiments 3 and 4



The punishment conditions were dummy coded (0: No punishment; 1: Punishment) All values represent unstandardised coefficients. *p < .05; **p < .01; ***p < .001. E3 = Experiment 3 (N = 159); E4 = Experiment 4 (N = 376).

Figure 3 Indirect effect of punishment on points returned to the third-party punisher via belief in the just world, deservingness, schadenfreude, and attitudes in Experiment 4



The punishment conditions were dummy coded (0: No punishment; 1: Punishment) All values represent unstandardised coefficients. *p < .05; **p < .01; ***p < .001

Footnotes

[1] In Experiments 1-3, we also included self-report measures of a range of emotions (Angry, Regretful, Jealous, Contemptuous, Satisfied, Happy, Guilty, Sad, Proud), followed by measures of impressions of other players (e.g. liking) and outcomes of the games (e.g. deservingness). These variables were included after each game partly to disguise the research's focus on schadenfreude and to avoid alerting participants to the fact that the outcomes for the players were pre-programmed, and partly to explore whether other emotions might influence the experience of schadenfreude after punishment. Findings that were both relevant to the research focus on Game 2 and were consistent across experiments were: firstly, anger having relatively consistent significant or near-significant positive associations with schadenfreude (Experiment 1: $r(39) = .30$, $p = .061$; Experiment 2: $r(77) = .27$, $p = .016$; Experiment 3: $r(80) = .26$, $p = .019$; Experiment 4: $p > .10$) and secondly, participants' emotions tended to decrease across time. Exploratory analyses of these emotions are included in the Supplementary Materials.

[2] Only 9 participants sent fewer than 40 points without this prompt, and all of these participants sent 40 points following the prompt. Analyses conducted on samples excluding participants who saw this prompt yielded similar results to those based on the entire sample.

[3] In response to an anonymous reviewer's comment, we performed an ANCOVA controlling for deservingness scores and rated satisfaction (from the emotion items described in footnote 1) to explore whether schadenfreude is simply a feeling of deservingness and satisfaction in response to justice concerns. Results found that the effect of punishment on schadenfreude remained significant, $F(1, 74) = 9.28$, $p = .003$, $\eta^2 = .11$, suggesting that

increased schadenfreude following punishment did not simply reflect increased deservingness and satisfaction, supporting a conceptual distinction between these constructs.

[4] Analyses including country as an IV revealed no significant effects of this variable. For the sake of clarity and economy, this IV was omitted from the analyses reported here.

[5] Nineteen participants reported suspicions about whether they were playing against other human players, 15 responded incorrectly to a memory check about the number of returned points, 1 transferred no points in game 1, and 4 used identification credentials that were identical to those of participants from a related previous study.

[6] As in Experiment 1, the effect of punishment on schadenfreude remained significant after controlling for deservingness and satisfaction, $F(1, 161) = 5.84$, $p = .017$, $\eta^2 = .04$

[7] Sixty-five participants failed to answer at least one attention check question correctly, 52 failed to complete the questionnaire, 23 reported suspicions concerning whether other players were real human beings, and 2 failed to transfer points to the trustee in game 1. Our dropout rate (17.28%) was lower than those (30% - 50%) reported in Zhou and Fishbach's study (2016) while our exclusion rate (21.59%) was within the range (2% - 52%) reported Thomas and Clifford's (2017) review.

[8] As in Experiments 1 and 2, the effect of punishment on schadenfreude remained significant after controlling for deservingness and satisfaction, $F(1, 153) = 15.05$, $p < .001$.

[9] A principal axis factor analysis with oblimin rotation was performed on the BJW and schadenfreude items. Both the eigenvalues (>1) and scree plot suggested a two-factor solution, explaining 62% of the variance. The first factor consisted of the seven BJW items and the second factor consisted of the four schadenfreude items. The two factors correlated at 0.49 and no item crossed-loaded onto the other factor above 0.30.

[10] Following Voorhees et al.'s (2016)'s recommendation, we used the overlapping confidence intervals approach (cf. Anderson & Gerbing, 1988) to show that the confidence intervals around the correlations did not include 1.0 (see Table 2), indicating the absence of serious discriminant validity issues.