What Stops Revenge Taking? Effects of Observed Emotional Reactions on Revenge Seeking

Andreas B. Eder & Vanessa Mitschke
Julius-Maximilians-University of Würzburg
Mario Gollwitzer
Ludwig-Maximilians-University of Munich

In press, Aggressive Behavior

Author Note
Andreas B. Eder, Department of Psychology; Vanessa Mitschke, Department of Psychology; Mario Gollwitzer, Department of Psychology.

This research was supported by grant Ed201/7-1 of the German Research Foundation (DFG) to AE. The funding agency had no role in the study design, collection, analysis or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication.

Preregistration documents, experiment files, and raw data can be accessed at https://osf.io/d7eb8/. A preprint of an earlier version is available at https://osf.io/yw6sk/.

Correspondence concerning this article should be addressed to AE, Department of Psychology, University of Würzburg, Röntgenring 10, 97070 Würzburg, Germany. Email: andreas.eder@uni-wuerzburg.de
WHAT STOPS REVENGE TAKING?

Abstract

What reaction stops revenge taking? Four experiments (total $N=191$) examined this question where the victim of an interpersonal transgression could observe the offender’s reaction (anger, sadness, pain, or calm) to a retributive noise punishment. We compared the punishment intensity selected by the participant before and after seeing the offender’s reaction. Seeing the opponent in pain reduced subsequent punishment most strongly, while displays of sadness and verbal indications of suffering had no appeasing effect. Expression of anger about a retributive punishment did not increase revenge seeking relative to a calm reaction, even when the anger response was disambiguated as being angry with the punisher. It is concluded that the expression of pain is the most effective emotional display for the reduction of retaliatory aggression. The findings are discussed in the light of recent research on reactive aggression and retributive justice.

*Keywords:* emotion display; aggression; retaliation;
What Stops Revenge Taking? Effects of Observed Emotional Reactions on Revenge Seeking

Victims of interpersonal transgressions often seek revenge by punishing the offender or making him suffer. Pertinent theorizing in psychology has focused on what exactly motivates victims to seek revenge (e.g., Yoshimura & Boon, 2018), whether and when taking revenge has hedonic benefits for the victim (e.g., Carlsmith, Wilson, & Gilbert, 2008; Eadeh, Peak, & Lambert, 2017; Gollwitzer, Meder, & Schmitt, 2011), and whether revenge-seeking inhibits or facilitates a victim’s willingness to forgive (e.g., Strelan & van Prooijen, 2013; Wenzel & Okimoto, 2014). In most of these studies, taking revenge was a one-shot behavior in which the avenger was unaware of how the offender (i.e., the target of revenge) reacted towards being punished. In real-life interactions, however, such reactions are often immediately visible to the avenger, and they may arguably have an impact on the avenger’s behavior, for instance, whether they continue or stop punishing the offender. To date, it is unknown which kind of reaction from the offender appeases the avenger most. This is a relevant question: addressing it contributes to a better understanding of the escalation of vengeful interactions and it can help to give practical advice on how to behave appropriately in such interactions.

Whenever punishment occurs face-to-face between a victim/avenger and the offender/target of revenge, it is highly likely that the offender’s reaction to the punishment (henceforth referred to as “target feedback”) affects the course of events in a vengeful episode. Research suggests a reciprocity norm that the quantity and quality of the revenge should be approximately proportional to the amount of harm implied in the original offense (Tripp & Bies, 1997). Consequently, the avenger will monitor the opponent for signs of inflicted harm, and should continue to aggress until an “appropriate” or “desired” level of harm was reached. This recursive
process can be understood as a closed feedback loop, in which the intended harm level is set as a reference level and the perception of the opponent’s state is the controlled variable (Carver & Scheier, 1982). Opponent reactions can be physical (e.g., bleeding), affective (e.g., moans), and social (e.g., begging the avenger to stop). The avenger should stop retaliating if the opponent shows the desired reaction.

Theories on revenge-seeking proposed different hypotheses about what opponent reaction appeases the avenger. According to the comparative suffering hypothesis, the offense caused an affective imbalance between the offender and the victim, and revenge-seeking will stop when the offender has suffered in a comparable way (Frijda, 1994). Hence, expression of suffering should be a particularly potent signal to the avenger to stop with punitive action. Complementary to comparative suffering, the avenger may also wish to teach the offender a lesson that his prior offense was condemnable and that a punishment is deserved (Miller, 2001). According to this understanding hypothesis, the avenger is appeased when the offender signals insight that the revenge was taken against him because and in virtue of their prior unfair behavior. Supportive of this hypothesis, several studies found that victims of injustice felt most satisfied with the outcome of their vengeful reaction when the original offender expressed understanding of the retribution (e.g., Funk, McGeer, & Gollwitzer, 2014; Gollwitzer & Denzler, 2009; Gollwitzer et al., 2011). In these studies, however, avengers did not see the target actually suffer; they merely received a written statement from the target in which they expressed (vs. did not express) understanding for the victim’s vengeful reaction as a response to their prior offense. Hence, it is unclear what nonverbal expression of the opponent is most appropriate for a de-escalation in vengeful interactions.
Displays of negative emotions are of particular relevance to this research question because they have a social significance in addition to the expression of suffering or annoyance (Keltner & Haidt, 1999; Shariff & Tracy, 2011). They communicate information about one’s feeling state, behavioral intentions, and requests for behavioral adaptations (Horstmann, 2003). As such, they can serve as incentives or deterrents for other individuals’ social behavior (Oosterhof & Todorov, 2008). For instance, expressions of anger could signal to another person that her behavior has violated a socially accepted standard and that behavioral adjustment is needed (Averill, 1983). In a vengeful interaction, expressing anger about a retribution is likely interpreted as disapproval of the retributive action, and consequently as a lack of understanding. The avenger, who is in control of the situation, could then desire an even harsher retribution to teach the adversary a lesson. Supportive of this hypothesis, a study showed that individuals with high power demand more compensation from angry than calm adversaries in a negotiation situation for which an anger response was inappropriate (Van Kleef & Côté, 2007). According to this model, a person will retaliate when she has high power over the situation and deems the expression of anger inappropriate to the situation at hand. Expression of sadness, by contrast, is most typically a signal of appeasement and communicates a request for help (Hackenbracht & Tamir, 2010; Hasson, 2009). In vengeful interactions, opponents’ displays of sadness could decrease revenge seeking by inducing a greater concern for the antagonist’s welfare in the observer. Supportive of this hypothesis, a study showed that negotiators conceded more to interaction partners who expressed sadness, especially if they felt responsible for the other’s feeling (Sinaceur et al., 2015). Thus, one could hypothesize contrary effects of sad and angry expressions on revenge seeking for interactions in which the person has high power and can risk further escalation of aggression.
Expressions of pain are expected to decrease revenge seeking according to the comparative-suffering hypothesis. Supportive of this hypothesis, early studies found that knowing that the target of aggression is in pain reduces further aggressions against this target (e.g., Geen, 1970). However, other research found the opposite: inflicting pain on others increased the likelihood of further aggressive acts (e.g., Sebastian, 1978; see also Bushman, 2002). According to the graduation hypothesis, the initial hedonic pleasure of inflicting pain can increase a desire to continue inflicting it (Wright & Hensley, 2003). Thus, expressions of pain can either reduce or exacerbate aggressive tendencies, and it is not clear what effect they unfold in vengeful interactions.

The present research aimed to clarify the (important and non-trivial) question which opponent response to a retaliatory punishment is most likely to reduce revenge taking.

**The present research**

In the experiments reported below, retaliatory aggression was provoked using a modified variant of the Taylor (1967) aggression paradigm. Specifically, participants played several rounds of a competitive reaction time game against a fictitious opponent and were punished by the opponent with an annoying noise blast if they lost the game. After a few lost games, they were given an opportunity to retaliate and could observe the opponent’s reaction to their (retributive) noise punishment. Emotional reactions were displays of pain, anger, and sadness. A calm expression was added for control. We were interested how participants will adjust the intensity of punishment in the next trial depending on the opponent reaction they have viewed in the previous trial. It should be noted that people often have difficulties to infer the feeling states of other people from observed facial displays (Zhou et al., 2017), that means, they could misinterpret the opponents’ expressions. Therefore, we also conducted experiments in which anger and pain
displays were combined with explicit indicators of anger feelings and/or suffering. These indicators also served to disambiguate the opponent response.

Study 1

Method

Participants

For each study, we planned to analyze data from a minimum of \( n=40 \) participants to detect an effect of \( d_z \geq .40 \) with acceptable statistical power \( (1-\beta=.80) \) and alpha set to .05. For Experiment 1, 62 volunteers were recruited via a departmental subject pool software. The analysis was run with \( n=44 \) (34 female, \( M_{age}=26.9 \) years, \( SD_{age}=7.9 \)) after exclusion of 18 participants according to our pre-registered criteria (see Data preparation below). The study protocols were approved by the ethics committee of the Department of Psychology, University of Würzburg (reference no. 2015-08).

Apparatus and material

A 3s long recording of white noise was used for noise punishment. Its intensity (max = 75dB) was varied in 5dB steps corresponding to each volume level (1-5). The opponent response was displayed in video clips showing an angry, sad, pain, or calm reaction to the sound blast (3s without audio). We selected 4 anger videos, 4 sadness, and 8 pain videos based on emotion ratings of 180 video clips in a pilot rating study. Raters \( (N=289) \) judged the emotionality of the observed reaction on self-assessment manikin scales (pleasantness, arousal, dominance) and expressions of anger, sadness, pain, disgust, and fear on unipolar scales (see the supplement for a documentation). The models (only males) expressing anger or sadness also provided videos with pain and calm displays. In total, participants viewed 108 videos (4 anger, 4 sad, 4 pain, 96 neutral) in a session. A static
picture of each model with a neutral expression was used for the introduction of the opponent. Materials are available at https://osf.io/d7eb8/.

**Procedure**

Participants were told that they would play online a competitive RT game with an ostensible participant located at another university. The “loser” in a game was punished with an annoying sound blast, and the participant could select the loudness of the sound blast (from 1=low to 5=very intense) delivered to the opponent. Aggression was indexed by the selection of volume levels on winning trials.

Before the game, the maximum volume of sound blast was adjusted individually based on the subjective maximally tolerable noise. Participants played games against several (fictitious) opponents who were introduced with a photograph (1000ms) at the start of each round. Each game round had 5 trials. Figure 1 shows the sequence of events in a trial. A trial started with the participant’s selection of the intensity of the sound blast that would be administered to the opponent on a winning trial. Then, a red circle appeared for a random time interval (500-800ms) for preparation. Participants were instructed to press the left mouse button as quickly as possible after the red circle has turned into green. A time limit of 1000ms was given for the response and the game was repeated in the case of a timeout. Upon response registration, bogus feedback on the winner was displayed with an indication of the selected intensity of the noise blast. Participants were informed that they were assigned to a condition in which they could observe the opponent during the sound blast in a live video transmission, while the opponent could not. On a winning trial, the participant hence saw the opponent’s reaction to the sound blast selected by her. On a losing trial, the participant heard the sound blast allegedly selected by the opponent. The next trial was initiated after 50ms.
A session had 48 game rounds (blocks), with each block containing 5 trials. In 16 (provocation) blocks, the fictitious opponent consistently selected very intense sound blasts (4-5) for punishment. In 32 (no-provocation) blocks, the opponent selected very low intensities (1-2). In the crucial experimental blocks, the participant lost the first three games and won the fourth and fifth game (see Fig. 1). This 3/2 loss/win streak was implemented in the 16 provocation blocks and in 8 non-provocation blocks. The remaining 24 blocks had other loss/win streaks (4 blocks: 1/4; 12 blocks: 2/3; 8 blocks: 4/1) that were intermixed to disguise the experimental blocks. In the provocation blocks, the opponent’s reaction on the first winning trial was emotional (angry, sad, painful) or calm (neutral). The reaction on the second winning trial was always calm. In no-provocation blocks, the opponent always reacted calmly. The participant played two game rounds against a single opponent: one in the first half of the blocks and a second round in the second half. If the opponent reacted with anger or sadness in the first round, he was calm or in pain in the second round, or vice versa. The assignment of the models to the counterbalanced conditions was random.
Figure 1. Competitive RT game against a (fictitious) opponent. A game round consisted of five trials (horizontal axis). The sequence of events in a trial is shown on the vertical axis. In the crucial blocks, the participant lost the first three games and won the last two games. Upon winning, participants could watch the opponent’s reaction during the sound blast in a “live” video transmission. Participants were (not) provoked with the opponent’s consistent selection of intense (low) sound blasts on losing trials (1-3), and they had an opportunity to retaliate on winning trials (4-5). Effects of target feedback on retaliatory aggression were indexed by volume adjustments from Trial 4 to Trial 5.

After each game round, participants rated feelings of pleasantness, dominance, and arousal using self-assessment manikin (SAM) scales (Bradley & Lang, 1994); justice satisfaction and
WHAT STOPS REVENGE TAKING?

deservingness (4 items; Cronbach’s α = .84) on 5-point scales adapted from Funk et al. (2014); and 6 items taken from the Aggressive Motives Scale (Anderson & Murphy, 2003) (see the supplement for the items). At the end, participants completed a validated German version of the Trait Aggression Questionnaire (von Collani & Werner, 2005) and were probed for suspicions about the purpose of the study.

**Data Preparation**

Our investigation of a moderation of retaliatory aggression by target feedback required that retaliation was successfully provoked in the first place. Therefore, we included only those data sets in our analyses that indicated retributive action by the participant. This was assessed by comparing the volume levels selected for the second and third trials in the provocation blocks with the volume levels in corresponding trials of no-provocation blocks. If the mean intensity was numerically higher in the provocation blocks, the data set was included in the analyses. This selection rule was pre-registered (https://osf.io/d7eb8/), and dropouts were immediately replaced during data collection to achieve our preregistered sample size (minimum n=40).

The dependent variable of main interest was the adjustment of punishment intensity after having viewed the opponent’s reaction to the sound punishment. Therefore, the intensity selected for the fourth trial in provocation blocks was subtracted from the intensity selected for the fifth trial, with negative scores indexing a reduction in aggressive punishments. Originally, we planned to analyze the difference scores using a repeated-measures analysis of variance (ANOVA) with opponent reaction (anger, sadness, pain, calm) as the within-subjects factor (see our preregistration documents at https://osf.io/d7eb8/). However, on reviewer suggestion, we switched to a multilevel analysis of the difference scores using a restricted maximum-likelihood linear mixed model (calculated with the GAMLj module for jamovi [version 1.0.7]); Galluci, 2019). The opponent
reaction (anger, sadness, pain, calm) was a fixed component, and the subject and movie intercepts were entered as random coefficients in the model. Fixed effects parameters estimates were tested for significance with the level of significance set at $p < .05$ corrected for multiple testing using the Holm method. Multilevel modeling capitalizes on a large number of trials and can control for error variation induced by two, or more, random factors (here: subjects and videos), which can improve the power of the statistical significance test (Chester, 2019; Judd et al., 2017). The reader is referred to the supplementary information file for a report of the (preregistered) analyses with repeated-measures ANOVAs. In addition, figures display mean scores of aggression data as a function of the conditions for convenient interpretation of the results.

**Results**

In line with our preselection rule, participants selected higher volumes in the provocation blocks ($M= 2.77$, $SD= 1.15$) than in the no-provocation blocks ($M= 1.77$, $SD= 0.82$), $t(43)=6.67, p < .001, d_z= 1.0$. In addition, participants felt less pleasant ($Ms[SDs]= 3.49 [0.66]$ vs. $3.71 [0.73]$), $t(43)=2.71, p = .009, d_z= 0.40$; more aroused ($Ms[SDs]= 2.23 [0.78]$ vs. $1.80 [0.77]$), $t(43)=6.03, p < .001, d_z= 0.90$; and less dominant ($Ms[SDs]= 3.40 [1.03]$ vs. $3.62 [0.95]$) in these blocks, $t(43)=3.62, p = .001, d_z= 0.54$; These differences indicate that the provocation was effective. Analyses of justice satisfaction, anger motives, and trait aggressiveness are reported in the supplemental information file.

In the omnibus test, the fixed effect of opponent reaction on the difference scores was significant, $F(3,20.1) = 7.02, p = .002$. Figure 2 shows that expression of pain reduced punishment most strongly. Volume levels were significantly reduced following displays of pain relative to calm expressions ($B =-0.51$, $SE = 0.11$), $t(33.01) = 4.45, p < .001$; anger displays ($B =-0.35$, $SE = 0.12$), $t(16.19) = 2.99, p = .043$; and expressions of sadness ($B =-0.34$, $SE = 0.12$), $t(16.19) = 2.94, p$
Intensity of punishment was also reduced after displays of anger ($B = -0.16, SE = 0.12$) and sadness ($B = -0.17, SE = 0.12$); however, these reductions were of comparable magnitude, $t(9.89) = 0.05, p = .963$, and they were not different from baseline, with $t(16.9) = 1.37, p = .231$, and $t(16.9) = 1.42, p = .172$, respectively.

**Figure 2.** Volume adjustment of noise punishments from Trial 4 to Trial 5 in Study 1 as a function of the opponent’s reaction to the noise punishment. Error bars show the 95% confidence interval. Conditions having common letter subscripts are significantly different at the .05 level corrected for multiple comparisons.

**Study 2**

In Study 1, expressions of pain reduced retaliation most strongly, while punishment after anger and sadness displays did not differ from baseline with calm reactions. A possible explanation is that participants did not interpret the opponent’s anger reaction as being angry about the
retaliation. In Experiment 2, we therefore disambiguated the reference of the anger response with an explicit indicator of how angry the fictitious opponent was with the participant after a punishment.

**Method**

**Participants**

The final sample comprised $n=46$ (38 female, $M_{age}=23.6$, $SD_{age}=3.8$) following the exclusion of 13 participants in line with our pre-registered criteria.

**Apparatus, stimuli, and procedure**

Setup and procedure were identical with Experiment 1 with the change that a 5-point anger item (“How angry are you with your opponent?”) was additionally included. On losing trials, the participant rated her anger feelings on this scale after the punishment; on winning trials, participants saw the anger rating of the fictitious opponent. In provocation blocks, the opponent’s anger rating was 5 (= very angry) after an anger response in the video, and 1 (= not at all) after the other videos. In no-provocation blocks, the fictitious opponent indicated no irritation (anger ratings with ‘1’ and ‘2’). We also had an anger item at the end of each game round after the SAM ratings that asked how angry they were with the opponent in this game round. Questionnaires of justice satisfaction; aggressive motives; trait aggressiveness were removed.

**Results**

In line with our preregistered selection rule, participants selected higher volumes in the provocation blocks ($M=3.37$, $SD=1.17$) than in the no-provocation blocks ($M=2.45$, $SD=0.98$), $t(45)=7.97$, $p<.001$, $d_{z}=1.18$. In provocation blocks, participants felt less pleasant ($M_{s/SDs}=[3.85/0.78]$ vs. $4.14/0.59$), $t(45)=3.78$, $p<.001$, $d_{z}=0.55$; more aroused ($M_{s/SDs}=[2.36/0.96]$ vs. $1.78/0.71$), $t(45)=5.37$, $p<.001$, $d_{z}=0.79$; and less dominant ($M_{s/SDs}=[3.48/1.04]$ vs. $3.72$
WHAT STOPS REVENGE TAKING?

[0.98]), \( t(45)=4.00, p < .001, d_{z}= 0.59. \) Participants were also more angry after noise punishment in these blocks \((M= 3.42, SD= 1.27)\) relative to no-provocation blocks \((M= 1.19, SD= 0.31)\), \( t(45)=11.86, p < .001, d_{z}= 1.75. \) Thus, provocation was effective.

In the multilevel model, the effect of opponent reaction on intra-individual differences in punishment was significant, \( F(3,687) = 5.67, p < .001. \) As shown in Fig. 3, pain displays again reduced punishments most strongly. Volume levels were not significantly reduced relative to calm reactions \((B = -0.24, SE= 0.12), t(36.8) = -2.04, p = .195, \) and sadness expressing no irritation \((B = -0.16, SE= 0.12), t(16.51) = -1.31, p = .414, \) but relative to anger expressions \((B = -0.48, SE= 0.12), t(16.51) = -4.03, p = .005. \) Reductions after displays of sadness were not different from those after anger \((B = -0.33, SE= 0.12), t(16.51) = -2.72, p = .115, \) or calm expressions \((B = -0.09, SE= 0.12), t(16.51) = -0.73, p = .479. \) Punishment intensity was least reduced after displays of anger with explicit indication of irritation feelings, albeit the difference to the baseline condition with calm reaction was not significant \((B = 0.24, SE= 0.12), t(16.51) = 1.99, p = .195. \)
Figure 3. Volume adjustment of noise punishments from Trial 4 to Trial 5 as a function of the opponent’s reaction to the noise punishment in Study 2. Error bars show the 95% confidence interval. Conditions having common letter subscripts are significantly different at the .05 level corrected for multiple comparisons.

The explicit feedback from the opponent that he was (not) irritated by the participant’s retributive punishment appears to have primarily affected the interpretation of sad and calm displays, while it enhanced the interpretation of the anger display as a hostile reaction. This interpretation is also supported by analyses of the participants’ anger ratings after a game round with provocations. A multilevel model with opponent reaction as fixed factor and subject and movie intercepts as random factors showed a clear effect of the opponent reaction on the anger ratings, $F(3, 14.6) = 10.1, p < .001$. Inspection of the means revealed that anger feelings were most intense after a game round with angry opponents ($M = 2.94, SD = 1.18$), while they were more moderate when the opponent reacted with sadness ($M = 2.46, SD = 0.95$), pain ($M = 2.53, SD = 1.10$), or calmly ($M = 2.67, SD = 1.00$).

Studies 3A and 3B

Study 2 suggests that revenge seeking is most reduced when the opponent expressed suffering and least when he expressed anger about the participant’s retributive action. In Experiments 3A and 3B, we combined opponent expressions of anger and pain with explicit indications of suffering and irritation, respectively. If knowledge that the opponent has suffered by the retaliation reduces revenge seeking, then anger displays with explicit indication of suffering should reduce subsequent punishments more than anger displays with indication of no suffering (Experiment 3A). If knowledge that the opponent was angry by the punishment increases revenge seeking, then pain displays with indication of anger feelings should reduce punishment less than
pain displays with explicit indication of no irritation (Experiment 3B). With these arrangements, we thus could find out what inference from the opponent reaction is more important for the regulation of revenge taking: the inference based on the opponent’s nonverbal behavior or the inference based on the verbal feedback from the opponent.

**Method**

**Participants**

Study 3A had $n=49$ (38 female, $M_{age}=23.2$, $SD_{age}=4.2$) after exclusion of 14 participants and Study 3B had $n=52$ (41 female, $M_{age}=23.3$, $SD_{age}=3.7$) after exclusion of 12 participants.

**Apparatus, stimuli, and procedure**

**Study 3A.** In addition to an explicit indication of irritation (as in Study 2), the fictitious opponent now also indicated suffering on a 5-point pain item (“How painful was the noise blast?”; 1 = not at all, 5 = very angry). On a losing trial, the participant rated her own feelings of pain; on a winning trial, she saw the opponent’s rating. Anger displays were paired with indicators of (a) either high irritation and low pain, (b) or high irritation and high pain. Pain displays were paired with indication of low irritation and high pain; calm displays with low irritation and low pain indication. The videos showing sadness were replaced with 4 more anger and pain videos (for details see the supplement). All other study details were identical with Study 2.

**Study 3B.** Procedure was the same with the major change that that the pain displays were paired with indicators of (a) either high irritation and high pain or (b) low irritation and high pain. Anger displays were paired with indication of high irritation and low pain.

**Results**

**Study 3A:** Participants selected higher volumes in the provocation blocks ($M=3.30$, $SD=1.19$) than in the no-provocation blocks ($M=2.50$, $SD=1.00$), $t(48)=7.51$, $p<.001$, $d_z=1.07$. In
addition, they felt less pleasant ($M$s/$SD$s = 3.50 [0.73] vs. 3.72 [0.79]), $t(48) = 2.75$, $p = .008$, $d_z = 0.39$; more aroused ($M$s/$SD$s = 2.41 [0.73] vs. 1.96 [0.68]), $t(48) = 4.74$, $p < .001$, $d_z = 0.67$; and less in control ($M$s/$SD$s = 3.31 [0.87] vs. 3.53 [0.88]), $t(48) = 2.56$, $p = .013$, $d_z = 0.36$. Anger feelings after punishment were higher in the provocation ($M = 3.50$; $SD = 1.16$) than in the no-provocation blocks ($M = 1.38$, $SD = 0.54$), $t(48) = 11.82$, $p < .001$, $d_z = 1.69$. These differences confirm that participants felt provoked.

In the multilevel model, the omnibus effect of the fixed factor opponent reaction (anger with indication of no suffering, anger with indication of suffering, pain, calm) was not significant, $F(3, 14.2) = 2.98$, $p = .067$. As shown in Fig. 4 (left panel), pain displays significantly lowered volume levels relative to calm reactions ($B = -0.40$, $SE = 0.14$), $t(28.1) = -2.87$, $p = .047$. In contrast, anger displays with explicit feedback of suffering produced no significant decrease relative to the baseline condition ($B = -0.13$, $SE = 0.15$), $t(16.2) = -0.85$, $p = 1.00$. Anger expressions with indication of no suffering also produced no difference ($B = -0.10$, $SE = 0.15$), $t(16.2) = -0.67$, $p = 1.00$. Notably, explicit feedback of (no) suffering did not influence the effects of anger displays on subsequent punishments ($B = -0.03$, $SE = 0.16$), $t(11.5) = -0.16$, $p = 1.00$. The effect of pain displays was not significantly different from the effects of anger displays with indication of suffering ($B = -0.30$, $SE = 0.15$), $t(16.2) = -2.02$, $p = .304$, and without suffering ($B = -0.27$, $SE = 0.15$), $t(16.2) = -1.84$, $p = .335$.

Anger ratings after a game round with provocations were analyzed with a multilevel model with opponent reaction as fixed factor and subject and movie intercepts as random factors. The omnibus test showed a significant effect of opponent reaction, $F(3, 16.7) = 6.30$, $p = .005$. Inspection of the means revealed that angry opponents with explicit indication of suffering ($M = \ldots$)
WHAT STOPS REVENGE TAKING?

3.14, $SD = 1.10$) and no suffering ($M = 3.01, SD = 1.08$) made participants more angry than calm opponents ($M = 2.78, SD = 1.05$) and opponents in pain ($M = 2.70, SD = 1.00$).

![Figure 4](image)

*Figure 4.* Adjustment of noise punishments from Trial 4 to Trial 5 as a function of the opponent’s reaction to the noise punishment in Studies 3A (left panel) and 3B (right panel). Error bars show the 95% confidence interval. Conditions having common letter subscripts are significantly different at the .05 level corrected for multiple comparisons.

**Study 3B:** Volume levels selected by the participant were higher in the provocation blocks ($M= 3.03, SD= 1.18$) than in the no-provocation blocks ($M= 2.28, SD= 0.97$), $t(51)=8.41, p < .001$, $d_z= 1.17$. Feelings after provocation were less pleasant ($Ms/SDs= 3.26 [0.79] \text{ vs. } 3.63 [0.85]$), $t(51)=3.64, p = .001$, $d_z= 0.50$; more arousing ($Ms/SDs= 2.37 [0.85] \text{ vs. } 1.76 [0.71]$), $t(51)=6.82, p < .001$, $d_z= 0.94$; and less dominant ($Ms/SDs= 2.89 [0.91] \text{ vs. } 3.25 [1.08]$), $t(51)=3.76, p < .001$, $d_z= 0.52$. In addition, participants were more angry with the opponent in provocation ($M= 3.72, SD= 1.09$) relative to non-provocation blocks ($M= 1.24, SD= 0.36$), $t(51)=15.28, p < .001$, $d_z= 2.11$.

In the omnibus test, the effect of *opponent reaction* (pain with indication of no irritation, pain with indication of irritation, anger, calm) was significant, $F(3, 777) = 3.97, p = .008$. As shown
in Fig. 4 (right panel), pain displays without feedback of irritation reduced punishments relative to calm expressions ($B = -0.41, SE = 0.13$), $t(16.42) = -3.16, p = .036$. Pain expressions with explicit feedback from the opponent that he was irritated about the punishment did not produce a significant reduction in punishment intensity relative to baseline ($B = -0.26, SE = 0.13$), $t(35.15) = -2.02, p = .206$. A direct comparison of pain displays with feedback of (no) irritation revealed no significant difference between both conditions ($B = -0.15, SE = 0.13$), $t(16.42) = -1.14, p = .579$. Expression of anger did not influence punishment intensities relative to calm displays ($B = -0.09, SE = 0.13$), $t(16.42) = -0.66, p = .579$, or pain displays with explicit indication of irritation ($B = 0.18, SE = 0.13$), $t(16.42) = 1.36, p = .579$. Latter condition was also not different from baseline with calm reactions ($B = -0.26, SE = 0.13$), $t(35.15) = -2.02, p = .206$.

The effect of the opponents’ reaction on anger ratings after a provocative game round was significant in the omnibus test of the multilevel model, $F(3, 776) = 3.44, p = .016$. Participants were angrier after games in which opponents expressed anger ($M = 2.98, SD = 1.08$) or pain and irritation ($M = 2.93, SD = 1.08$) relative to game rounds with calm opponents ($M = 2.81, SD = 1.04$) and opponents expressing pain and no irritation ($M = 2.71, SD = 0.90$).

**Mini meta-analysis**

Effects of emotional expressions on revenge seeking were meta-analyzed using fixed effects in which the mean effect size (mean difference) was weighted by sample size. For ease of analyses, we computed Pearson’s correlation coefficients for each effect size using the formula described by Goh, Hall, and Rosenthal (2016, p. 541). A positive correlation coefficient indexed more reduction in punishment intensity relative to the condition with calm expressions, while a negative correlation indexed less reduction of revenge seeking. Correlation coefficients were then Fisher’s z-transformed for analyses (performed with the MAJOR module for jamovi [version
1.0.7]; Hamilton, 2019) and converted back to Pearson correlation for presentation. Figure 5 shows a forest plot of the results.

For the meta-analysis of pain, effects of pain displays (without indication of irritation in Study 3B) relative to calm displays were selected ($k = 4$). The meta-analysis showed a highly significant effect, mean $r = .39$, $Z = 5.43$, $p < .001$, two-tailed. Thus, expression of pain reduced subsequent punishments substantially relative to calm expressions.

For the meta-analysis of anger, effects of anger displays (without indication of suffering in Study 3A) relative to calm displays were selected ($k = 4$). The overall effect was not significant, mean $r = .02$, $Z = 0.40$, $p = .691$, two-tailed.

For the meta-analysis of sadness, effects of sadness displays relative to calm displays were selected ($k = 2$). Overall, the effect was not significant, mean $r = .16$, $Z = 1.42$, $p = .156$, two-tailed.
Figure 5. Forest plots of effect sizes (raw correlation coefficients) with a summary estimate (mean correlation) separately for each viewing condition (pain, anger, sadness). Squares are proportional to the weights used in the meta-analysis; lines show the 95% confidence interval. Positive effect sizes indicate more reduction of punishment relative to the viewing condition with calm opponent reactions.
General Discussion

Four experiments (total n=191) investigated whether opponents’ emotional reactions to a retaliatory punishment affects the avenger’s willingness to provide further punishments. The results were clear-cut: Compared to expressions of anger, sadness, and calmness, expressions of pain reduced subsequent punishment most strongly. This reduction was remarkably consistent across experiments (Study 1: -0.57; Study 2: -0.64; Study 3A: -0.57; Study 3B: -0.57), and it was only slightly attenuated when the opponent explicitly expressed irritation about the punitive action (Study 3B: -0.42). By contrast, punishment was reduced to the least extent following displays of anger (Study 1: -0.22; Study 3A: -0.27; Study 3B: -0.25), especially when the anger response was explicitly disambiguated as being angry with the participant (Study 2: -0.16). Adding explicit feedback of suffering to the anger display made no difference (Study 3A: -0.29), showing that simply knowing that the target has suffered is not sufficient to make revenge seeking stop. Opponents’ expressions of sadness had no appeasing effect on the victim that would be different from calm expressions. In short, nonverbal displays of pain stopped revenge seeking, while other emotional displays and verbal indications of suffering had no effect.

The reduction of aggression after having viewed the target in pain is in line with the hypothesis that the avenger has achieved its goal when the offender has suffered to the same extent as the victim had suffered (Frijda, 1994). Importantly, our findings corroborate the notion that “comparative suffering” should not be confused with an “eye-for-an-eye” principle (or a “tit-for-tat” rule): Tit-for-tat means reciprocating the offender’s action (Axelrod & Hamilton, 1981). However, the noise levels selected by participants following provocation were, on average, substantially lower than the noise levels selected by the fictitious opponent for provocation (for descriptive data see Tables S1 and S2 in the supplement). In addition, participants did not endorse
the statement that they intended to pay back the opponent for the noise levels he set (see the ratings in Table S4 in the supplement). Nevertheless, participants stopped seeking revenge after displays of pain. This implies that it was not the motivation to reciprocate the offense (sound blast) itself but, rather, the motivation to reciprocate the suffering it caused, which affected their retaliatory response.

Although the decrease of punishment levels following pain displays is consistent with the notion of ‘comparative suffering,’ it can also be interpreted alternatively. One alternative interpretation is that seeing the offender in pain elicited some form of compassion or empathic concern in the avengers, and that this negative affective state made them stop seeking revenge. While some findings suggest that observing another’s pain automatically elicits an empathetic response in the observer (Singer & Lamm, 2009), there is also evidence that empathetic responses are reduced (predominantly in males) when observing an unfair person in pain (Singer et al., 2006). Thus, it is unclear whether an empathic response is plausible in conditions of provocation. Another interpretation could be that seeing the offender suffer made avengers feel guilty, and that this guilt has made them stop taking revenge (Haidt, 2003). Interview studies indeed suggest that avengers often feel guilt or shame after revenge taking (e.g., Boon, Alibhai, & Deveau, 2011; see also Carlsmith et al., 2008; Eadeh et al., 2017). While we did not ask our participants directly for feelings of guilt and shame, their ratings of justice satisfaction and deservingness did not vary as a function of the opponent response (see the supplement). In short, we cannot clearly tell on the basis of the present data whether a norm of reciprocity, empathy for pain, or moral emotions can explain our findings better. Future research should therefore clarify what processes were triggered by seeing the target in pain, and how they interact with personality characteristics of the avenger.
Results were not in line with the understanding hypothesis. According to this hypothesis, revenge-seeking should have decreased after the opponent has signaled understanding that he was punished for his prior offense (Miller, 2001). Reacting with anger to a punishment should have signaled disapproval of the retaliation—and hence a lack of understanding of the retributive action. As a consequence, participants should have increased the intensity of punishment after having observed the anger reaction, which was not found in the present studies. Participants were expected to intensify the punishment when they deem the opponent’s expression of anger inappropriate to the situation at hand and can risk further escalation (Van Kleef & Côté, 2007). It is possible that these conditions were only partly met in the present task. Another possibility is that the comparison with calm expression was not fair, because the avenger could have also interpreted a calm reaction as a lack of insight. While this objection could work for Study 1, it is less plausible for the other studies with explicit feedback from the opponent that he was not angry about the retaliation. Clearly, more research is needed on what nonverbal behaviors signal understanding to avengers, and in which conditions they will weaken retributive action tendencies.

The present research also has limitations. One limitation is that opponents were only males, while most participants were females. It is possible that gender differences in the expression and/or perception of emotions have influenced the results (Kret & De Gelder, 2012). For instance, people commonly believe that males can endure more physical pain than females (Wise et al., 2002), and gender roles affect the acceptance of physical aggression as a means for retaliation (Bettencourt & Miller, 1996). Women also report more fear in confrontation with males in anger-producing circumstances, and males are more intimidating in such situations than females (Brody et al., 1995). Men who express neutral and angry emotions are also rated as higher in dominance in comparison with men expressing sadness, whereas females showing a neutral expression are perceived as less
dominant (Hareli et al., 2009). Consequently, it is possible that the levels of revenge motivation will be different when the subject/aggressor is a male and/or the target of aggression a female. Women also show stronger empathetic responses to pain displays (Christov-Moore et al., 2014), and they more readily accept overt expression of pain as an appropriate behavior (Nayak et al., 2000). Hence, it is an open research question whether expressions of pain will be similarly effective if viewed by a male. For a systematic investigation of gender effects, future studies could vary the sex of the participants and the opponents observed during the task.

Another limitation of the present research is that trait aggressiveness of our student sample and the intensity of (provoked) punishments were generally low (for descriptives see Table S5 in the supplement). Although we obtained clear evidence that the provocation was effective, students were presumably inhibited to use a physical means for retaliation. It would be interesting to scrutinize the generalizability of our results to more diverse populations (with regard to aggression norms) and/or to other forms of retributions. Participants could have also feared retaliation in the present studies because they played two game rounds against a single opponent. In this case, however, punishment should have been most inhibited with angry opponents expressing irritation, which was clearly not the case. Therefore, we believe that fear of retaliation was not a strong factor in the present research. For further inquiry, future studies could include explicit measures of retaliation fear and vary the number of game rounds played against an (angry) opponent.

To summarize, the present results suggest that avengers will cease seeking revenge when they see the target of their revenge in pain. It is evident that this perception is complex and potentially biased by a host of factors, such as personality characteristics and social norms. This complexity also explains why revenge-seeking often does not stop when the target suffers (Stillwell et al., 2008). Aggression reduction interventions hence should not only attempt to
WHAT STOPS REVENGE TAKING?

sensitize aggressors to the suffering of the victim; they should also educate potential targets of aggression to express their suffering clearly.
References

Anderson, C. A., & Murphy, C. R. (2003). Violent video games and aggressive behavior in young women. *Aggressive Behavior, 29*(5), 423–429. https://doi.org/10.1002/ab.10042

Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist, 38*(11), 1145–1160. https://doi.org/10.1037/0003-066X.38.11.1145

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*(4489), 1390–1396. https://doi.org/10.1126/science.7466396

Bettencourt, B. A., & Miller, N. (1996). Gender differences in aggression as a function of provocation: A meta-analysis. *Psychological Bulletin, 119*(3), 422–447. https://doi.org/10.1037/0033-2909.119.3.422

Boon, S. D., Alibhai, A. M., & Deveau, V. L. (2011). Reflections on the costs and benefits of exacting revenge in romantic relationships. *Canadian Journal of Behavioural Science, 43*(2), 128–137. https://doi.org/10.1037/a0022367

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Brody, L. R., Lovas, G. S., & Hay, D. H. (1995). Gender differences in anger and fear as a function of situational context. *Sex Roles, 32*(1), 47–78. https://doi.org/10.1007/BF01544757

Bushman, B. J. (2002). Does venting anger feed or extinguish the flame? Catharsis, rumination, distraction, anger, and aggressive responding. *Personality and Social Psychology Bulletin, 28*(6), 724–731. https://doi.org/10.1177/0146167202289002
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

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin, 92*(1), 111–135. https://doi.org/10.1037/0033-2909.92.1.111

Chester, D. S. (2019). Beyond the aggregate score: Using multilevel modeling to examine trajectories of laboratory-measured aggression. *Aggressive Behavior, 45*(5), 498–506. https://doi.org/10.1002/ab.21837

Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience & Biobehavioral Reviews, 46*, 604–627. https://doi.org/10.1016/j.neubiorev.2014.09.001

Eadeh, F. R., Peak, S. A., & Lambert, A. J. (2017). The bittersweet taste of revenge: On the negative and positive consequences of retaliation. *Journal of Experimental Social Psychology, 68*, 27–39. https://doi.org/10.1016/j.jesp.2016.04.007

Frijda, N. H. (1994). The lex talionis: On vengeance. In S. H. M. van Goozen, N. E. Van de Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 263–289). Lawrence Erlbaum Associates, Inc.

Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the message: Punishment is satisfying if the transgressor responds to its communicative intent. *Personality and Social Psychology Bulletin, 40*(8), 986–997. https://doi.org/10.1177/0146167214533130

Galluci, M. (2019). *GAMLj: General analyses for linear models [jamovi module]*. (Version 2.0.0) [Computer software]. https://gamlj.github.io/
WHAT STOPS REVENGE TAKING?

Geen, R. G. (1970). Perceived suffering of the victim as an inhibitor of attack-induced aggression. *The Journal of Social Psychology, 81*(2), 208–215. https://doi.org/10.1080/00224545.1970.9922442

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass, 10*(10), 535–549. https://doi.org/10.1111/spc3.12267

Gollwitzer, M., & Denzler, M. (2009). What makes revenge sweet: Seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology, 45*(4), 840–844. https://doi.org/10.1016/j.jesp.2009.03.001

Gollwitzer, M., Meder, M., & Schmitt, M. (2011). What gives victims satisfaction when they seek revenge? *European Journal of Social Psychology, 41*(3), 364–374. https://doi.org/10.1002/ejsp.782

Hackenbracht, J., & Tamir, M. (2010). Preferences for sadness when eliciting help: Instrumental motives in sadness regulation. *Motivation and Emotion, 34*(3), 306–315. https://doi.org/10.1007/s11031-010-9180-y

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

Hamilton, K. (2019). MAJOR: Meta analysis JamOvi R [jamovi module]. (Version 1.0.0) [Computer software]. https://github.com/kylehamilton/MAJOR

Hareli, S., Shomrat, N., & Hess, U. (2009). Emotional versus neutral expressions and perceptions of social dominance and submissiveness. *Emotion, 9*(3), 378–384. https://doi.org/10.1037/a0015958
Hasson, O. (2009). Emotional tears as biological signals. *Evolutionary Psychology, 7*(3), 363–370. https://doi.org/10.1177/147470490900700302

Horstmann, G. (2003). What do facial expressions convey: Feeling states, behavioral intentions, or actions requests? *Emotion, 3*(2), 150–166.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68*(1), 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion, 13*(5), 505–521.

Kret, M. E., & De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia, 50*(7), 1211–1221. https://doi.org/10.1016/j.neuropsychologia.2011.12.022

Miller, D. T. (2001). Disrespect and the experience of injustice. *Annual Review of Psychology, 52*(1), 527–553. https://doi.org/10.1146/annurev.psych.52.1.527

Nayak, S., Shiflett, S. C., Eshun, S., & Levine, F. M. (2000). Culture and gender effects in pain beliefs and the prediction of pain tolerance. *Cross-Cultural Research, 34*(2), 135–151. https://doi.org/10.1177/106939710003400203

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Sebastian, R. J. (1978). Immediate and delayed effects of victim suffering on the attacker’s aggression. *Journal of Research in Personality, 12*(3), 312–328. https://doi.org/10.1016/0092-6566(78)90058-2
Shariff, A. F., & Tracy, J. L. (2011). What are emotion expressions for? Current Directions in Psychological Science, 20(6), 395–399. https://doi.org/10.1177/0963721411424739

Sinaceur, M., Kopelman, S., Vasiljevic, D., & Haag, C. (2015). Weep and get more: When and why sadness expression is effective in negotiations. The Journal of Applied Psychology, 100(6), 1847–1871. https://doi.org/10.1037/a0038783

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. Annals of the New York Academy of Sciences, 1156(1), 81–96. https://doi.org/10.1111/j.1749-6632.2009.04418.x

Singer, T., Seymour, B., O’Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. Nature, 439(7075), 466–469. https://doi.org/10.1038/nature04271

Stillwell, A. M., Baumeister, R. F., & Del Priore, R. E. (2008). We’re all victims here: Toward a psychology of revenge. Basic and Applied Social Psychology, 30(3), 253–263. https://doi.org/10.1080/01973530802375094

Strelan, P., & van Prooijen, J.-W. van. (2013). Retribution and forgiveness: The healing effects of punishing for just deserts. European Journal of Social Psychology, 43(6), 544–553. https://doi.org/10.1002/ejsp.1964

Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. Journal of Personality, 35(2), 297–310. https://doi.org/10.1111/j.1467-6494.1967.tb01430.x

Tripp, T. M., & Bies, R. J. (1997). What’s good about revenge? The avenger’s perspective. In R. J. Lewicki, R. J. Bies, & B. H. Sheppard (Eds.), Research on negotiation in organizations, Vol. 6 (pp. 145–160). Elsevier Science/JAI Press.
Van Kleef, G. A., & Côté, S. (2007). Expressing anger in conflict: When it helps and when it hurts. *Journal of Applied Psychology, 92*(6), 1557–1569. http://dx.doi.org/10.1037/0021-9010.92.6.1557

von Collani, G., & Werner, R. (2005). Self-related and motivational constructs as determinants of aggression: An analysis and validation of a German version of the Buss–Perry Aggression Questionnaire. *Personality and Individual Differences, 38*(7), 1631–1643. https://doi.org/10.1016/j.paid.2004.09.027

Wenzel, M., & Okimoto, T. G. (2014). On the relationship between justice and forgiveness: Are all forms of justice made equal? *British Journal of Social Psychology, 53*(3), 463–483. https://doi.org/10.1111/bjso.12040

Wise, E. A., Price, D. D., Myers, C. D., Heft, M. W., & Robinson, M. E. (2002). Gender role expectations of pain: Relationship to experimental pain perception. *Pain, 96*(3), 335–342. https://doi.org/10.1016/S0304-3959(01)00473-0

Wright, J., & Hensley, C. (2003). From animal cruelty to serial murder: Applying the graduation hypothesis. *International Journal of Offender Therapy and Comparative Criminology, 47*(1), 71–88. https://doi.org/10.1177/0306624X02239276

Yoshimura, S. M., & Boon, S. D. (2018). *Communicating revenge in interpersonal relationships*. Lexington Books.

Zhou, H., Majka, E. A., & Epley, N. (2017). Inferring perspective versus getting perspective: Underestimating the value of being in another person’s shoes. *Psychological Science, 28*(4), 482–493. https://doi.org/10.1177/0956797616687124
Author contributions

AE developed the study concept. All authors contributed to the study design. Testing and data collection were performed by VM, who also performed the data analysis under the supervision of AE. AE drafted the manuscript, and the co-authors provided critical revisions. All authors approved the final version of the manuscript for submission.
Supplemental Online Material

Descriptive statistics of noise punishments

Table S1

Mean volume levels (with SD) in each condition before and after the opponent reaction.

|                  | Sadness | Anger | Pain (provoked) | Pain (unprovoked) |
|------------------|---------|-------|-----------------|-------------------|
|                  | Before  | After | Before          | After             | Before  | After | Before  | After | Before  | After |
| Study 1          | 3.11    | 2.88  | 3.23            | 3.01              | 3.10    | 3.04  | 1.70    | 1.75  |
|                  | (1.32)  | (1.39)| (1.33)          | (1.38)            | (1.23)  | (1.22)| (0.88)  | (0.95)|
| Study 2          | 3.69    | 3.21  | 3.69            | 3.53              | 3.77    | 3.38  | 1.95    | 1.98  |
|                  | (1.34)  | (1.33)| (1.37)          | (1.23)            | (1.30)  | (1.42)| (1.10)  | (1.10)|

Table S2

Mean volume levels (with SD) in each condition before and after the opponent reaction.

|                  | Anger with pain/Pain with anger | Anger | Pain (provoked) | Pain (unprovoked) |
|------------------|---------------------------------|-------|-----------------|-------------------|
|                  | Before  | After | Before          | After             | Before  | After | Before  | After | Before  | After |
| Study 3A         | 3.61    | 3.32  | 3.63            | 3.37              | 3.48    | 3.31  | 2.05    | 2.05  |
|                  | (1.27)  | (1.33)| (1.34)          | (1.30)            | (1.39)  | (1.22)| (1.11)  | (1.09)|
| Study 3B         | 3.37    | 2.95  | 3.45            | 3.21              | 3.29    | 3.13  | 1.80    | 1.95  |
|                  | (1.35)  | (1.31)| (1.42)          | (1.33)            | (1.28)  | (1.32)| (1.05)  | (1.18)|

Pre-registered analysis of Study 1

The dependent variable of main interest was the adjustment of punishment intensity after having viewed the opponent’s reaction to the sound punishment. Therefore, the intensity selected for the fourth trial in provocation blocks was subtracted from the intensity selected for the fifth trial, with negative scores indexing a reduction in aggressive punishments. Difference scores were analyzed using a repeated-measures analysis of variance (ANOVA) with opponent reaction (anger, sadness, pain, calm) as the within-subjects factor. Follow-up t-tests compared difference scores after
emotional displays (anger, sadness, pain) and with that of the baseline condition (calm display) with the level of significance set at $p<.05$ (two-tailed) after Sidák correction for multiple comparisons.

The opponent reaction had a significant effect on the intensity of subsequent punishments, $F(3,129)=7.23, p < .001$, $\eta^2_p = .144$. Follow-up comparisons showed that expression of pain reduced punishment most strongly relative to the other displays, while displays of anger and sadness had no effect (see Table S3).

Table S3

| Pairwise comparison       | Mean difference† [95% CI] | $p^b$ | $d_z$ |
|---------------------------|--------------------------|-------|-------|
| Pain-anger                | $-0.347^*$ [-0.65, -0.05] | .016  | 0.48  |
| Pain-sadness              | $-0.341^*$ [-0.61, -0.07] | .007  | 0.52  |
| Pain-calm                 | $-0.506^*$ [-0.88, -0.13] | .003  | 0.57  |
| Anger-sadness             | 0.006 [-0.24, 0.25]       | 1     | < 0.01|
| Anger-calm                | 0.159 [-0.48, 0.16]       | .681  | 0.21  |
| Sadness-calm              | 0.165 [-0.49, 0.16]       | .680  | 0.21  |

† Negative values index more punishment reduction in the first viewing condition.
* The mean difference is significant at the .05 level.

$^b$ Adjusted for multiple comparisons after Sidák.

**Analyses of justice satisfaction and deservingness.** Participants rated four items (I am satisfied with the outcome of the game; In the end, everybody got what they deserved; I feel satisfied now; I feel like justice has been restored) on a scale from 1 (= ‘not at all’) to 5 (= ‘extremely’). Sum
scores were analyzed with opponent reaction as factor. The Greenhouse-Geisser corrected ANOVA showed no significant effect, $F(2.5, 107.5) = 1.88, p = .148, \eta^2_p = .042$. Justice satisfaction was rated highest after having watched the opponent suffer ($M = 10.51, SD = 3.4$) compared to angry ($M = 10.23, SD = 3.3$), sad ($M = 10.20, SD = 3.1$), and calm reactions ($M = 10.06, SD = 3.1$).

**Analyses of aggressive motives.** Ratings of the aggressive motives in provocation blocks are shown in Table S4. Separate ANOVAs for each motive with opponent reaction (anger, pain, sadness, calm) as factor showed no significant effect (largest $F = 1.36, ps > .26$).

**Table S4**

*Mean ratings (with SD) of aggressive motives in the provocation blocks.*

| Motive Description                                                                 | Rating (SD)  |
|-------------------------------------------------------------------------------------|--------------|
| I wanted to impair my opponent’s performance in order to win more.                  | 1.51 (0.9)   |
| I wanted to control my opponent’s level or responses.                               | 1.83 (1.1)   |
| I wanted to drive my opponent mad.                                                  | 1.62 (0.9)   |
| I wanted to hurt my opponent.                                                       | 1.91 (1.3)   |
| I wanted to pay back my opponent for the noise levels he set.                        | 1.63 (1.0)   |
| I wanted to blast him harder than he blasted me.                                     | 2.54 (1.3)   |

*Note.* Motives were rated after each game round on unipolar scales from 1 = not at all to 5 = a lot.

**Analyses of trait aggressiveness.** Scores of two participants were lost due to a coding error. Table S5 shows summary statistics of the remaining sample ($n=42$). The four subscales were highly intercorrelated (Cronbach’s $\alpha = .823$). Linear regression of a difference score indexing the intensity of provocation on the total scores of trait aggressiveness yielded no significant effect ($\beta = .497, t < 1$). Linear regression analyses of difference scores indexing volume adjustments following the opponent reaction (with negative scores indicating a decrease in punishment intensity) revealed a significant equation for displays of anger, $F(1, 40) = 4.45, p = .041$, with an $R^2$ of .10. Punishments
following angry responses increased $B = 0.009$ for each unit of trait aggressiveness. Regression equations for the other conditions were not significant (largest $F = 3.26, ps > .070$).

Table S5

*Trait aggressiveness scores.*

| Trait          | Mean (SD)  | Min-Max |
|----------------|------------|---------|
| Physical aggression | 21.7 (8.0) | 9-49    |
| Verbal aggression    | 17.2 (5.4) | 12-38   |
| Anger              | 22.8 (7.0) | 12-38   |
| Hostility          | 22.9 (9.3) | 10-51   |
| Total score        | 84.6 (24.4)| 42-147  |

**Pre-registered analysis of Study 2**

Opponent reactions had a significant effect on intra-individual differences in punishment, $F(3,135)=5.13, p=.002, \eta^2_p = .102$. As shown in Fig. 3 in the article, pain displays again reduced punishment most strongly. Follow-up tests revealed that the reduction of noise punishment was significant only in the comparison with anger displays and not in other comparisons (see Table S6).

Table S6

*Pairwise comparisons of volume adjustments after opponent viewing in Study 2*

| Pairwise comparison | Mean difference† [95% CI] | $p^b$ | $d_z$ |
|---------------------|---------------------------|-------|-------|
| Pain-anger          | -0.484‡ [-0.85, -0.11]    | .005  | 0.53  |
| Pain-sadness        | -0.158 [-0.50, 0.19]      | .762  | 0.19  |
| Pain-calm           | -0.245 [-0.60, 0.11]      | .323  | 0.28  |
| Anger-sadness       | 0.326 [0, 0.65]           | .05   | 0.41  |
| Anger-calm          | 0.239 [-0.10, 0.58]       | .307  | 0.28  |
| Sadness-calm        | -0.087 [-0.44, 0.26]      | .984  | 0.10  |
**WHAT STOPS REVENGE TAKING?**

† Negative values index more punishment reduction in the first viewing condition.
* The mean difference is significant at the .05 level.
Adjusted for multiple comparisons after Sidák.

**Pre-registered analysis of Study 3A**

The ANOVA of the volume difference scores with *opponent reaction* as factor (anger with indication of no suffering, anger with indication of suffering, pain, calm) was significant, $F(3,144)=3.52$, $p=.017$, $\eta^2_p = .068$. In follow-up comparisons, only the contrast of pain displays with calm displays was significant (see Table S7).

Table S7

*Pairwise comparisons of volume adjustments after opponent viewing in Study 3A*

| Pairwise comparison                  | Mean difference† [95% CI] | $p^b$  | $d_c$ |
|-------------------------------------|---------------------------|--------|-------|
| Pain-anger without suffering        | -0.301                    | .093   | 0.36  |
| Pain-anger with suffering           | -0.276                    | .114   | 0.34  |
| Pain-calm                           | -0.398*                   | .05    | 0.39  |
| Anger without suffering-anger with suffering | 0.026                    | 1      | 0.03  |
| Anger without suffering-anger       | -0.097                    | .982   | 0.10  |
| Anger without suffering-calm        | -0.122                    | .937   | 0.13  |
| Pain-anger with suffering           | 0.026                     | 1      | 0.03  |
| Pain-anger with suffering           | -0.276                    | .114   | 0.34  |
| Pain-calm                           | -0.398*                   | .05    | 0.39  |
| Anger without suffering-anger with suffering | 0.026                    | 1      | 0.03  |
| Anger without suffering-anger       | -0.097                    | .982   | 0.10  |
| Anger without suffering-calm        | -0.122                    | .937   | 0.13  |

† Negative values index more punishment reduction in the first viewing condition.
* The mean difference is significant at the $p < .05$ level.
Adjusted for multiple comparisons after Sidák.
Pre-registered analysis of Study 3B

The ANOVA with *opponent reaction* as factor (pain with indication of no irritation, pain with indication of irritation, anger, calm) was significant, \(F(3,153)=3.53, p=.016, \eta^2_p = .065\). In follow-up comparisons, pain displays without feedback of irritation reduced punishments relative to calm expressions significantly (see Table S8).

Table S8

*Pairwise comparisons of volume adjustments after opponent viewing in Study 3B*

| Pairwise comparison        | Mean difference\(^\dagger\) [95% CI] | \(p^b\)  | \(d_z\)  |
|---------------------------|--------------------------------------|---------|---------|
| Pain without irritation-pain with irritation | -0.149 [-0.57, 0.27] | .913 | 0.14 |
| Pain without irritation-anger | -0.327 [-0.66, 0.01] | .060 | 0.37 |
| Pain without irritation-calm | -0.413\(^*\) [-0.78, -0.05] | .020 | 0.43x |
| Pain with irritation-anger | -0.178 [-0.58, 0.22] | .792 | 0.17 |
| Pain with irritation-calm | -0.264 [-0.64, 0.11] | .311 | 0.27 |
| Anger-calm                 | -0.087 [-0.46, 0.29] | .990 | 0.09 |

\(^\dagger\) Negative values index more punishment reduction in the first viewing condition.

\(^*\) The mean difference is significant at the \(p < .05\) level.

\(^b\) Adjusted for multiple comparisons after Sidák.
Selection of video clips

Models for the video clips were 26 male students. The model was seated in front of a laptop against a grey backdrop and wore headphones that were also used during the competitive reaction time game. Videos were recorded with 1920x1080 pixels using a webcam (Logitech Europe S.A., Lausanne, Switzerland). Victim responses were instructed verbally by a researcher. Each model produced four different reactions (pain, anger, sadness, calm) in response to a simulated noise blast. The video was cut to 3 sec length after recording.

Two naïve coders selected the three most unambiguous expressions of each emotion category (pain, anger, sadness) and for each model. Videos of six models were excluded due to insufficient expressiveness. Emotional reactions of the remaining 20 models were rated by N = 289 in an online study recruited via the Prolific platform (https://prolific.ac/). A total of 180 videos was divided into 3 subsets, each containing 3 emotional reactions from each model. Videos showing a calm reaction were not included. Participants rated the emotionality of the reaction (“How the person in the video might feel?”) on self-assessment manikin scales (pleasantness: 1=unhappy, 5=happy; arousal: 1=calm, 5=excited; dominance: 1=being controlled, 5=in-control) and expressions of anger, sadness, pain, disgust, and fear (“Please rate how intensely the person in the video might feel [emotion]”) on unipolar scales ranging from 1=“not at all” to 5=“extremely.” Subjects also rated the authenticity (“How genuine is the person's expression?”) on a scale ranging from 1=“not at all” to 5=“extremely.”

Means for the intensity ratings of pain, anger and sadness were calculated and 90% CIs were constructed around each mean. Videos were selected that had a high intensity rating on the relevant emotion category and no overlapping CIs with the other emotion dimensions. For Study 1 and 2, we selected four different models for the anger and sadness condition each (see Table S9).
Each of the anger and sadness expressions were paired with one pain expression and two neutral expressions from that model (32 videos in total). For Studies 3A and 3B, we selected an additional set of 4 models showing anger, and pain or calm expressions (see Table S10). Additional videos with calm reactions were portrayed by the remaining 18 models.

The video clips can be downloaded at https://osf.io/d7eb8/. The models signed an agreement to the publication of the video material in the internet. The use and distribution of the videos is regulated by the CC Attribution 4.0 International Public License.

Table S9

Emotional ratings of the victim reactions selected for Studies 1 and 2

| Video | Model no. | Anger | Sadness | Pain |
|-------|-----------|-------|---------|------|
|       |           | M (SD) | CI      | M (SD) | CI   | M (SD) | CI   |
| Anger reaction | 5       | 4.36 (1.01) | [4.18;4.55] | 1.91 (1.05) | [1.72;2.11] | 2.46 (1.25) | [2.23;2.70] |
|         | 6       | 2.75 (1.30) | [2.53;2.98] | 2.20 (1.00) | [2.02;2.37] | 1.76 (.91) | [1.60;1.92] |
|         | 20      | 2.79 (1.13) | [2.60;2.99] | 2.45 (1.05) | [2.26;2.63] | 1.86 (1.03) | [1.68;2.04] |
|         | 26      | 3.15 (1.26) | [2.92;3.39] | 1.58 (.76) | [1.43;1.72] | 1.60 (.88) | [1.44;1.76] |
| overall |         | 3.26 (.75) | [2.38;4.15] | 2.03 (.37) | [1.59;2.47] | 1.92 (.38) | [1.48;2.36] |
| Sadness reaction | 1       | 1.38 (.72) | [.126;1.51] | 3.83 (1.14) | [.363;4.02] | 2.75 (1.28) | [2.53;2.97] |
|         | 18      | 1.44 (.81) | [1.31;1.58] | 3.11 (1.17) | [.292;3.31] | 1.87 (1.08) | [1.69;2.05] |
|         | 22      | 1.38 (.68) | [1.25;1.50] | 3.43 (1.03) | [.323;3.62] | 2.01 (1.09) | [1.81;2.21] |
|         | 21      | 1.66 (.87) | [1.50;1.83] | 3.26 (1.13) | [.305;3.47] | 1.69 (.89) | [1.52;1.85] |
| overall |         | 1.47 (.14) | [1.31;1.62] | 3.41 (.31) | [.304;3.77] | 2.08 (.47) | [1.53;2.63] |
| Pain reaction | 1       | 1.61 (.94) | [1.45;1.76] | 1.78 (.99) | [1.61;1.94] | 3.09 (1.39) | [2.86;3.32] |
|         | 5       | 2.23 (1.07) | [2.03;2.42] | 2.19 (1.06) | [1.99;2.38] | 3.35 (1.09) | [3.15;3.55] |
|         | 6       | 1.55 (.84) | [1.39;1.71] | 2.41 (1.26) | [2.18;2.65] | 3.61 (1.36) | [3.40;3.87] |
|         | 18      | 1.48 (.86) | [1.33;1.62] | 1.75 (1.00) | [1.58;1.92] | 3.01 (1.25) | [2.80;3.22] |
|         | 20      | 2.16 (1.30) | [1.94;2.39] | 2.07 (1.11) | [1.87;2.26] | 3.30 (1.28) | [3.08;3.53] |
|         | 21      | 1.51 (.97) | [1.33;1.69] | 1.81 (1.09) | [1.61;2.02] | 2.76 (1.21) | [2.54;2.99] |
|         | 22      | 1.53 (.85) | [1.39;1.68] | 2.38 (1.22) | [2.17;2.59] | 3.20 (1.21) | [2.99;3.41] |
|         | 26      | 1.69 (.99) | [1.52;1.85] | 1.89 (1.00) | [1.72;2.06] | 2.83 (1.21) | [2.63;3.03] |
| overall |         | 1.72 (.30) | [1.52;1.92] | 2.03 (.27) | [1.85;2.21] | 3.14 (.29) | [2.96;3.33] |
### Table S10

*Emotional ratings of the victim reactions selected for Studies 3A and 3B*

| Video no. | Model no. | Anger        | Sadness       | Pain        |
|-----------|-----------|--------------|---------------|-------------|
|           |           | $M$ (SD)     | CI            | $M$ (SD)    | CI          | $M$ (SD)    | CI          |
| 5         | 1         | 4.36 (1.01)  | [4.18;4.55]   | 1.91 (1.05) | [1.72;2.11] | 2.46 (1.25) | [2.23;2.70] |
| 6         | 5         | 2.75 (1.30)  | [2.53;2.98]   | 2.20 (1.00) | [2.02;2.37] | 1.76 (.91)  | [1.60;1.92] |
| 20        | 2         | 2.79 (1.13)  | [2.60;2.99]   | 2.45 (1.05) | [2.26;2.63] | 1.86 (1.03) | [1.68;2.04] |
| 26        | 2         | 3.15 (1.26)  | [2.92;3.39]   | 1.58 (.76)  | [1.43;1.72] | 1.60 (.88)  | [1.44;1.76] |
| 2         | 3         | 2.71 (1.25)  | [2.50;2.92]   | 1.65 (.95)  | [1.49;1.81] | 1.57 (.96)  | [1.73;1.41] |
| 8         | 3         | 3.63 (1.28)  | [3.41;3.85]   | 1.58 (.93)  | [1.42;1.74] | 1.65 (.93)  | [1.49;1.81] |
| 18        | 1         | 2.83 (1.31)  | [2.58;3.07]   | 1.70 (.79)  | [1.55;1.85] | 1.76 (.98)  | [1.58;1.95] |
| 12        | 2         | 3.67 (1.23)  | [3.46;3.89]   | 2.88 (1.37) | [2.64;3.12] | 2.54 (1.29) | [2.32;2.77] |
|           | overall   | 3.24 (.60)   | [2.84;3.64]   | 1.99 (.48)  | [1.67;2.31] | 1.90 (.38)  | [1.64;2.16] |
|           | 1         | 1.61 (.94)   | [1.45;1.8]    | 1.78 (.99)  | [1.61;1.94] | 3.09 (1.39) | [2.86;3.32] |
|           | 5         | 2.23 (1.07)  | [2.03;2.42]   | 2.19 (1.06) | [1.99;2.38] | 3.35 (1.09) | [3.15;3.55] |
|           | 6         | 1.55 (.84)   | [1.40;1.71]   | 2.41 (1.26) | [2.18;2.65] | 3.61 (1.36) | [3.40;3.87] |
|           | 18        | 1.48 (.86)   | [1.3;1.62]    | 1.75 (1.00) | [1.58;1.98] | 3.01 (1.25) | [2.80;3.22] |
|           | 2         | 1.69 (1.01)  | [1.52;1.86]   | 2.26 (1.23) | [2.06;2.47] | 2.84 (1.18) | [3.04;2.64] |
|           | 3         | 1.52 (.90)   | [1.37;1.68]   | 1.64 (.90)  | [1.49;1.80] | 2.21 (1.16) | [2.01;2.41] |
|           | 8         | 1.66 (.94)   | [1.49;1.84]   | 1.89 (.90)  | [1.72;2.06] | 2.63 (1.10) | [2.42;2.83] |
|           | 12        | 2.15 (.98)   | [1.97;2.33]   | 2.28 (1.03) | [2.08;2.47] | 2.55 (1.10) | [2.35;2.76] |
|           | overall   | 1.83 (.30)   | [1.63;2.03]   | 2.08 (.26)  | [1.90;2.25] | 2.91 (.47)  | [2.60;3.23] |