



A Painful Message

Testing the Effects of Suffering and Understanding on Punishment Judgments

Eyal Aharoni^{1,3}, David Simpson³, Eddy Nahmias³, and Mario Gollwitzer⁴

¹Department of Psychology, Georgia State University, Atlanta, GA, USA

²Department of Philosophy, Georgia State University, Atlanta, GA, USA

³Neuroscience Institute, Georgia State University, Atlanta, GA, USA

⁴Department of Psychology, LMU, Munich, Germany



Abstract: [author, please supply an Abstract of around 150 words max.]

Keywords: [author, please supply up to five keywords]

Perhaps in no other social context are human beings more motivated to condone, support, and even celebrate the suffering of other people than when we punish violators of social norms. We accommodate and systematize the delivery of punishment across domains of social life, including interpersonal, legal, and international contexts. But punishing others is risky and costly, so why do we so readily engage in it? Research in social psychology has suggested that typical individuals espouse multiple punishment motives (Carlsmith & Darley, 2008; Gromet & Darley, 2009; Vidmar & Miller, 1980), including the desire for retribution (McFatter, 1982; Orth, 2004; Vidmar, 1974; Warr et al., 1983) and consequentialist aims like deterrence (Crockett et al., 2014; Ellsworth & Ross, 1983; Vuk et al., 2020), and rehabilitation (Gromet & Darley, 2009). However, when tested in rivalrous contexts, retribution may play a dominant role (Aharoni & Fridlund, 2012; Carlsmith, 2006; Carlsmith, Darley, & Robinson, 2002; Carlsmith, Monahan, & Evans, 2007; Darley, Carlsmith, & Robinson, 2000). Despite the support for this “retributivism hypothesis,” it remains unclear (a) what proximate goals retribution is designed to realize and (b) whether this pattern generalizes across offenses of different levels of seriousness and punisher perspectives (i.e., personal victim of the crime vs. impersonal). These open questions call for a conceptual replication of the retributivism hypothesis.

Research on the psychological mechanisms driving retribution has identified a shortlist of contenders, including the motivation to make perpetrators suffer in proportion to the harm as a means to re-establish a sense of justice achieved (see Berman, 2010; Frijda, 1994) and the motivation to

make perpetrators understand why they have been punished (see Duff, 2001, 2002; Miller, 2001). Thus, the “suffering hypothesis” would accord with strict deontological accounts of retribution, which seek to justify punishment on purely retrospective grounds, using principles of proportionality (Berman, 2010; Crombag et al., 2003; Gerber & Jackson, 2013; Moore, 1997), whereas the “understanding hypothesis” would be most consistent with expressive and communicative theories of punishment. These theories justify punishment as a means to express the victim’s and the community’s condemnation of perpetrators’ wrongdoing and to communicate to them that they violated a norm and must apologize and reform (Berman, 2010; Cushman et al. [author: please update], in press; Funk et al., 2014; Nahmias & Aharoni, 2017).

To the extent that these factors are viewed as predictive of the perpetrator’s future dangerousness, displays of suffering, understanding, and remorse might also be relevant to consequentialist theories of punishment. Indeed, the proximate motivations of real-world punishment judgments may commingle in complex and interesting ways that have been largely ignored in the literature. For this reason, the question of their relative and combined influence on lay punishment attitudes is valid and important in its own right, regardless of which of the classical punishment theories these constructs ultimately best support.

In an effort to test the relative contributions of suffering and understanding, Gollwitzer and colleagues (Gollwitzer & Denzler, 2009; Gollwitzer et al., 2011) conducted a set of punishment experiments using various measures of “goal-fulfillment” in response to some manipulated information

(such as introducing evidence of suffering or understanding). Together, their findings suggest that, while evidence of perpetrator suffering increased victims' sense of justice achieved to some degree, goal-fulfillment after punishing the perpetrator was most clearly observable when the perpetrator demonstrated an understanding of *why* he or she had been punished (Gollwitzer & Denzler, 2009; Gollwitzer et al., 2011). Thus, these studies provide solid empirical evidence for the "understanding hypothesis" but only weak and inconsistent evidence for the "suffering hypothesis."

Other studies using economic games have shown similar effects, finding, for instance, that anonymous participants are more likely to shift from a severe to moderate punishment strategy when they believe cheaters would be informed of why they have been penalized – namely because they treated their partner unfairly (Molnar et al., 2020). Support for the communicative function of punishment has even been found among young children, who elected to incur greater personal costs to punish wrongdoers who would be informed about the punishment than those who would not (Marshall et al., 2020; Twardawski & Hilbig, 2020). These studies are insightful because they help to identify the characteristic features of punishment goal fulfillment. The results of such studies thus contribute to theoretical models of punishment, but even more, they help to clarify the social utility of different punishment strategies and to predict when the delivery of a given punitive sanction will terminate, or instead, escalate.

Despite these scientific advances, open questions remain about the stability and generalizability of these effects. First, many of the studies described above did not independently manipulate the perpetrator's level of suffering. For instance, Gollwitzer and colleagues (2011 [author: citation ok now?]) tested the "suffering hypothesis" by comparing punishment conditions to a "fate" condition in which participants were informed that the ostensible perpetrator had "bad luck." Importantly, the amount of harm done to the perpetrator was allegedly equal between the punishment and the "fate" condition, and the perpetrator's actual suffering was not visible to participants. Only one recent study used a direct display of the perpetrator's suffering after punishment and varied it experimentally (Eder et al., 2020). In this study, which measured retaliatory aggression in response to provocations by an opponent on an ostensible winning streak, participants reduced their retaliatory aggressions when the opponent displayed a facial expression of pain (but not other emotions), suggesting empirical support for the "suffering hypothesis." It is unclear whether these differences between studies are attributable to differences in the task format, in the way suffering is defined, or some other factor.

Second, these tests of the suffering and understanding hypotheses have largely adopted a dyadic, two-party

structure (Eder et al., 2020; Funk et al., 2014; Gollwitzer & Denzler, 2009; Gollwitzer et al., 2011). In that format, the punisher, who also is the victim, has a vested interest in the delivery of punishment, making it distinct from impersonal sanctions, such as criminal sentencing, in which the punisher (e.g., a judge or jury) does not have direct, personal stakes in the punishment. From a traditional rational choice perspective, impersonal punishers should not engage in costly punishment, and yet research has shown that they regularly do so (Fehr & Fischbacher, 2004). Some research suggests that they are even more punitive than personal punishers, namely the victims of the crime (e.g., Zhou et al., 2017).

One recent vignette study (Bauer & Poama, 2020) examined, in an impersonal context, the independent effects of evidence of perpetrator suffering and moral change on perceptions that justice had been achieved by prison punishment. The moral change was operationally defined as a sincere apology and desire to help crime victims, implying that the perpetrator came to understand the gravity of his offense. Although evidence of suffering did increase participants' justice perceptions, evidence of moral change had a considerably larger effect. While such findings are illuminating, none of the previous studies has directly manipulated the personal versus impersonal perspective, so it is unclear to what extent punishment by impersonal punishers is similarly motivated by a desire that the perpetrator suffer *per se*, or to what extent their punitive motives can be satisfied by signals of understanding, even in the absence of suffering.

A third question about the existing punishment literature arises from the fact that many of these studies employ economic-style games, so they are limited to the study of relatively benign norm violations, like inequitably distributing small amounts of game money. As a result, it remains unclear whether their findings generalize to more serious types of violations, such as various types of criminal behavior. On the one hand, more serious norm violations should evoke a stronger desire for suffering, and perhaps greater demand for understanding and remorse. But on the other hand, for these more serious offenses, people may have higher demands for evidence that the perpetrators actually suffer and actually understand that what they did was wrong. In one study, for instance, researchers found that when a hypothetical rape was portrayed as more intentional, people attributed less remorse to the perpetrator (Kleinke et al., 1992). Presumably, the seriousness of the crime raised participants' thresholds for what constitutes genuine remorse. Because of the conflicting predictions on this matter, we make no predictions about moderating effects of offense seriousness on punishment behavior. However, any evidence that crime seriousness does or does not moderate the effects of understanding or suffering will be a valuable contribution to the punishment literature.

A fourth question is whether any effects of suffering or understanding on punishment goal fulfillment are influenced by self-reported support for conventional justifications for punishment (i.e., retribution, utility, rehabilitation, and communication). We make no a priori predictions about these relationships (for results, see Appendix B, available at <https://dx.doi.org/10.23668/psycharchives.5005>).

On a related point, existing research on the roles of suffering versus understanding typically frames these two motivations as independent, but more complex relationships are possible. For example, do people desire more punishment when suffering and understanding are both absent, or could the presence of understanding actually evoke more punishment, perhaps by implying greater criminal intent? This latter possibility suggests that researchers need to be careful about how to define understanding. If understanding means that the perpetrator fully understood the risk of harm even before the commission of the crime but did it anyway, we would not expect this kind of understanding to pacify punishers; if anything, it should increase punishment further. In order for understanding to achieve punishment goal fulfillment, it should be portrayed as a newfound realization that is a retrospective response to the crime. Similarly, a deep understanding does not just mean that perpetrators understand that their actions caused harm. It means that they believe that their actions were wrong and feel remorseful about the harm they caused (see Duff, 2001). In other words, the type of understanding that is most likely to satisfy our punitive goals is one that includes remorse as a part of its definition.

Defined in this way, and replicating previous research (Funk et al., 2014; Gollwitzer & Denzler, 2009; Gollwitzer et al., 2011; Molnar et al., 2020) we predict that punishment goal fulfillment will be greatest when perpetrators evidently suffered from the punishment *and* understand that what they did was wrong. Conversely, punishment goal fulfillment will be weakest when perpetrators show no signs of having suffered and/or having understood the wrongfulness of their actions. Such an interaction might suggest that people utilize suffering as a communicative device because they perceive it to be an effective way of inducing a deep understanding (Cushman et al., in press). In this way, evidence of suffering could serve as an indication that the perpetrator has internalized the message. Such an outcome would be consistent with hybrid theories of punishment, including evolutionary (e.g., Cushman, 2015) and communicative theories (e.g., Duff, 2001; Nahmias & Aharoni, 2017). These open questions about the relative and interactive roles of suffering and understanding raise a demand for a conceptual replication and extension of this important body of research.

The present project addressed these questions with a design that differs from previous research in several ways

but nevertheless represents a conceptual replication of this research. Our specific objective was to test the effects of perpetrator suffering and understanding in the expression of retributive punishment attitudes toward criminal offenders. The *understanding hypothesis*, as found in previous research (e.g., Funk et al., 2014; Gollwitzer & Denzler, 2009; Gollwitzer et al., 2011; Molnar et al., 2020), predicts that the signal that perpetrators understand why they have been punished (and regret it) will increase satisfaction with the punishment and reduce additional punishment recommendations, relative to the no understanding condition (a main effect of understanding). The *suffering hypothesis*, also receiving some support (e.g., Eder et al., 2020), predicts that the induction of suffering in the perpetrator will be sufficient to evoke satisfaction with the punishment and reduce additional punishment recommendations. We also expected that indicators of punishment goal fulfillment would be greater in the presence of both suffering and understanding than in any other combination of understanding and suffering (i.e., a synergistic interaction effect).

To test these hypotheses in a criminal punishment context, our study employed a contrastive vignette method rather than a behavioral laboratory task as reported in previous studies. Our study also uniquely included manipulations of crime seriousness and party perspective and a unique measure of support for common philosophical justifications for punishment in order to explore the potential influence of these variables on our hypothesized effects. However, core elements of those previous studies were preserved, namely the comparison of perpetrator suffering (present vs. absent) and perpetrator understanding (present vs. absent), and the inclusion of punishment and satisfaction measures to estimate punishment goal fulfillment (Gollwitzer & Denzler, 2009; Gollwitzer et al., 2011). In this way, our approach represents a conceptual replication of previous work on intuitive retributivism. All study methods were preregistered at PsychArchives.org. For the complete pre-registration protocol and study materials, see <http://dx.doi.org/10.23668/psycharchives.3160>.

Method

Sample Size Determination

As explained above, our crucial independent variables are understanding (absent vs. present) and suffering (absent vs. present), and we expected two main effects and interaction between the two. Previous research found medium to strong effects for the understanding hypothesis (Gollwitzer et al., 2011, Study 1: $d = 0.96$; Study 3: $d = 1.22$; Funk et al., 2014; Study 1: $d = 0.94$; Study 2a: $d = 1.17$; Study 2b: $d = 0.63$) as well as for the suffering hypothesis

(Eder et al., 2020: meta-analytic $r = 0.39$), so we assumed a medium-size effect for our two main effects as well. For the hypothesized interaction effect, determining the population effect was more difficult because no previous study has tested such an interaction effect before. Assuming the interaction effect is smaller than the main effects, we determined it to be small to medium in size.

As we explain below in more detail, we used a 2 (suffering present vs. absent) \times 2 (understanding present vs. absent) \times 2 (personal vs. impersonal party) \times 2 (crime seriousness low vs. high) \times 2 (vignette order) mixed factorial design with random assignment to conditions, where Suffering, Understanding, Party, and Order were manipulated between subjects and Crime Seriousness was varied within subjects. The data were analyzed using mixed linear modeling. Since determining the necessary sample size in a mixed linear model depends on a number of a priori assumptions (Westfall et al., 2014), which are difficult to make in our case, we based our sample size estimation on power analysis for conventional two-factorial ANOVA models. Assuming a small to medium-size two-way interaction effect (understanding \times suffering) of $f = .15$, $N = 472$ cases are needed to detect such an effect with $\alpha = .05$ and $1-\beta = .90$. Note that this sample size estimation is conservative, given that we used mixed linear modeling instead of conventional ANOVA. In other words, the actual power to detect a small- to medium-size interaction effect with this sample size is likely higher.

Participants

One thousand three hundred thirty-one adults consented to the survey. Participation was restricted by age (18+ years) and country (US) and approximately matched to US Census data on age, gender, and education. Three hundred forty-five participants were excluded for incomplete data; 11 for failing a multiple-choice attention check ("What are the colors of the American flag?"); 197 for failing to recognize the correct crime from a multiple-choice list; 217 were excluded for failing to recognize the assigned party perspective (personal or impersonal); 34 were excluded for indicating they would change their punishment recommendation but failing to do so, or vice versa; 13 for being more than two standard deviations above the mean in terms of duration, and 12 for taking less than five minutes. The remaining 514 cases were used for our analyses, which is a larger sample size than necessary (see above). The participants in this sample self-reported as: 46.5% male, 53.5% female; 8.2% Hispanic or Latino; 83.9% White/Caucasian, 8.2% Black or African American, 5.1% Asian, and 2.0% other/unknown/prefer to not answer (ethnic and racial categories were non-exclusive); and with a mean age of 48.6 ($SD = 17.5$).

Hypotheses

The Understanding Hypothesis (H1)

The signal that the perpetrator understands why he has been punished will increase satisfaction with the prospect of parole and reduce additional prison sentence recommendations relative to the signal that the perpetrator does not understand why he has been punished, regardless of whether the perpetrator suffered. Such an effect would constitute a main effect of understanding.

The Suffering Hypothesis (H2)

Evidence of suffering will increase satisfaction with the prospect of parole and reduce additional prison sentence recommendations compared to evidence that the perpetrator did not suffer, regardless of whether the perpetrator understands why he is being punished. This pattern would constitute a main effect of suffering.

The Understanding by Suffering Hypothesis (H3)

Indicators of punishment goal fulfillment will be greatest among the combined presence of suffering and understanding (a synergistic interaction effect). By contrast, goal fulfillment will be weaker if suffering, understanding, or both are absent.

Design

This experiment employed a contrastive vignette method, presenting text-based criminal case summaries that varied in systematic ways. The design was comprised of a 2 Suffering (present vs. absent) \times 2 Understanding (present vs. absent) \times 2 Party (personal vs. impersonal) \times 2 Crime Seriousness (low vs. high) \times 2 Vignette Order mixed factorial design with random assignment to conditions, where Suffering, Understanding, Party, and Order were manipulated between subjects and Crime Seriousness was varied within subjects. Varying crime types within-subjects enabled us to increase statistical sensitivity by using each subject as their own control. To offset the risk of order effects, the order of the two crime vignettes was reversely counterbalanced. This strategy has been employed successfully by other investigators in this research area (e.g., Darley et al., 2000).

The less serious crime was a second-degree theft (an employee stole a winning lottery ticket for \$5,000 that was purchased with money from a collective pool and intended for a group of ten), and the more serious crime was an aggravated robbery (a perpetrator robbed a bystander and injured them with a weapon). Theft-related crimes were selected to more closely approximate the economic-type norm violations that were used in previous studies testing the competing hypotheses (e.g., Gollwitzer et al., 2011). The two crime vignettes (impersonal version) were

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

pretested in an independent sample of university undergraduates to ensure that they differed in perceived seriousness ($p < .01$). Suffering was operationally defined as testimonial evidence from the perpetrator's jail therapist that he has already suffered greatly as a direct result of his initial time in jail of 10 weeks (e.g., job loss, mental and physical distress). The "suffering absent" conditions included testimony that he had not suffered during this jail time. The understanding was defined as testimonial evidence from the therapist that the perpetrator now regrets his actions, understands why they were wrong, and wishes to apologize for them. The "understanding absent" conditions included testimony that he does not believe his actions were wrong, does not regret them, and refuses to apologize. In keeping with the view that deep understanding requires remorse and apology, we included these two elements in our definition of understanding. The Party manipulation stipulated that the victim of the offense was the participant (personal) or an unrelated community member (impersonal). Manipulation checks confirmed that the perpetrators who were portrayed as having higher levels of understanding and suffering were in fact perceived as such (for details, see Appendix A, available at <https://dx.doi.org/10.23668/psycharchives.5005>)

Two dependent measures were designed to estimate punishment goal fulfillment. First, we administered a prison sentencing scale ("How long in prison should the defendant be incarcerated for this offense?" 0-5 years for the less serious crime; 0-10 years for the more serious crime). This measure was delivered both before and after the experimental induction, from which a pre-post difference score was derived (with larger decreases in recommended sentences indicating greater punishment goal fulfillment and larger increases indicating lack of punishment goal fulfillment). The purpose of this repeated measure was to observe attitude change directly within individuals rather than to merely infer it from random assignment to conditions. This approach also afforded greater statistical sensitivity by using each participant as his/her own control. However, this design also permits a fully-between-subjects analysis of post-manipulation sentencing scores by themselves.

Second, we administered a Likert-type scale measuring satisfaction with the prospect of immediate parole ("Suppose that after his 10-week jail time, [perpetrator's name] was granted immediate parole in the community instead of prison time. Further, suppose that he did not commit any further crimes during parole. How satisfied

or dissatisfied are you with parole as his only sentence?"), after the experimental induction. This use of the satisfaction measure was inspired by Gollwitzer and colleagues (2011) but departed somewhat from the pre-registration protocol. In the protocol, we proposed to include a baseline measure of punishment satisfaction in addition to the post-manipulation measure of satisfaction described here. However, the baseline measure of satisfaction was removed prior to data collection because of concerns that it was uninformative (i.e., most people would presumably be highly satisfied with their own baseline punishment recommendation). The remaining satisfaction measure refers to satisfaction with the prospect of early parole, which was delivered after the study manipulations only. The rationale for this measure is that the greater the punishment goal fulfillment, the more the participant would be satisfied by the prospect of early parole, which is generally regarded as less severe than the same period of time in prison. In Western countries like the US, framing the alternative as early parole tends to be more realistic than an unconditional discharge and encourages participants to make use of the full-scale range.

To test our confirmatory hypotheses, a linear mixed model was constructed using suffering and understanding as between-subjects factors, crime seriousness as a within-subjects factor, and sentencing change scores as the dependent measure.¹ The general form of this model is presented below:

$$y_{ij} = \gamma_{00} + \gamma_{01}S_j + \gamma_{02}U_j + \gamma_{03}C_{ij} + \gamma_{04}(S_j \times U_j) + \gamma_{05}(S_j \times C_{ij}) + \gamma_{06}(U_j \times C_{ij}) + \gamma_{07}(S_j \times U_j \times C_{ij}) + u_{0j} + e_{ij}, \quad (1)$$

where y_{ij} is the reported satisfaction value of participant j at measurement occasion i , S_j is a dummy variable representing perpetrator suffering (0 = absent, 1 = present), U_j is a dummy variable representing perpetrator understanding (0 = absent, 1 = present), C_{ij} is a dummy variable representing crime type (0 = low seriousness, 1 = high seriousness), $\gamma_{00} \dots \gamma_{07}$ are model parameters, u_{0j} represents the level-2 (i.e., between-participants) random intercept, and e_{ij} represents the level-1 (i.e., within-participants) residual.

In this form, all model parameters are in reference to the group defined by the intercept term γ_{00} . Model estimation was performed using the Restricted Maximum Likelihood method. The covariance type was unstructured (implying no assumptions that could be violated). Multiple

¹ To determine if participants' responses differed based on the order they received the vignettes (high or low seriousness crime first), we ran two independent-samples t -tests for unequal variances, which revealed no significant order effects on punishment recommendation changes, $t(118.529) = -1.559, p = .122$, or parole satisfaction ratings, $t(512) = -0.965, p = .336$, using each participant's high and low crime seriousness summary score. We thus opted not to enter the order into our formal hypothesis tests. [author: please include footnote into main text, if possible]

479 comparisons were adjusted using a Bonferroni correction.
 480 This analytic strategy represents a departure from our
 481 pre-registration protocol, which proposed to test our
 482 hypotheses using a Multivariate Analysis of Variance. This
 483 change was made because our study design was not fully
 484 crossed (every subject got one low and one high crime seri-
 485 ousness), so a linear mixed model was determined to be
 486 more appropriate for evaluating all planned comparisons.

487 Procedures

488 To ensure that our sample would emulate the broader adult
 489 population along basic demographic attributes, sampling
 490 was conducted by a professional sampling company
 491 (Respondi) and funded by PsychLab, a service of the
 492 Leibniz Institute for Psychology (ZPID). The sampling com-
 493 pany employed multi-source methods, including recruiting
 494 from a variety of Internet platforms such as advertisements
 495 on websites, email recruitment (to a preexisting pool), face-
 496 to-face recruitment, and telephone recruitment. Since the
 497 participants were recruited from multiple sources, the
 498 sampling company used algorithms to minimize source
 499 bias. One potential drawback of Internet samples is
 500 under-representing people with limitations on Internet
 501 access, such as the elderly and those with less than a high
 502 school education (see Smith, 2014). Such characteristics,
 503 thus, were monitored and strategically over-sampled (i.e.,
 504 stratified) or weighted until their proportion matched that
 505 of the broader US population.

506 The survey was delivered on the Qualtrics survey plat-
 507 form, which supports random assignment to between-
 508 subjects conditions. Individuals who responded to the
 509 survey invitation navigated (via hyperlink) to a survey con-
 510 sent form. Following consent, participants read the survey
 511 instructions and two case summary vignettes, describing a
 512 high and low seriousness crime from either a personal or
 513 impersonal perspective. The high and low seriousness
 514 vignettes were counterbalanced for order.

515 After each of the two vignettes, the following steps
 516 occurred: Participants were prompted to make a baseline
 517 prison punishment rating. Then, they were exposed to the
 518 experimental induction, namely the testimonial evidence
 519 of either or both understanding and suffering being present
 520 or absent (also counterbalanced for order). Next, partici-
 521 pants completed the two dependent measures collectively
 522 designed to estimate punishment goal fulfillment. First,
 523 they were asked if they wanted to change their initial
 524 sentence based on the new information, and if so, they
 525 completed the final prison sentencing scale; then, they
 526 responded to the parole satisfaction scale. (As expected,
 527 responses to these two dependent measures were

negatively correlated, $r = -.199$, $p < .001$. The correlation
 between the post-manipulation sentencing scores and
 satisfaction ratings was even stronger, $r = -.617$, $p <$
 $.001$.) Following the dependent measures, participants
 answered several manipulation check questions and a ques-
 tion about their memory for the crime case summaries.

Having completed this process for each crime vignette,
 participants were asked to complete a punishment justifica-
 tion ranking scale, a question about criminal justice system
 involvement, an attention check question, and standard
 demographic questions (see Materials section for further
 details). Last, participants were debriefed about the purpose
 of the study. All study procedures were approved by the
 host university's ethical review board, and participation
 was contingent upon informed consent.

Materials

All stimuli, including the crime vignettes, the text of the
 experimental induction, and primary dependent measures,
 were developed internally. (See pre-registration protocol
 for details.) The planned manipulation checks included
 Likert-type questions about whether the perpetrator fully
 understands that his actions were wrong, whether he genu-
 inely suffered as a result of his 10-week jail stay, attitudes
 about the early parole alternative (4 items, not including the
 primary satisfaction measure), and how harmful the crime
 was (-3 to $+3$). The memory check questions asked partici-
 pants to select from a multiple-choice list which crime the
 perpetrator committed and which party perspective they
 were asked to adopt.

A self-reported punishment justification scale, adapted
 from Bauer and Poama (2020) and Nadelhoffer and col-
 leagues (2013), asked participants to rank their endorse-
 ment of four distinct statements, counterbalanced for
 order, about the proper justification for criminal punish-
 ment. This scale was designed to capture the most salient
 features of the main punishment theories (retributive, con-
 sequentialist, and communicative). Although the individual
 statements are not necessarily unique to one punishment
 theory (e.g., the desire to "send a message to the offender"
 could be inspired by communicative, retributive, and/or
 consequentialists motivations), each statement was crafted
 to be more typical of one than the other.

One item asked participants to report any prior involve-
 ment in criminal trials for oneself or a family member
 (either as a defendant or an accuser). One attention check
 asked participants to answer a multiple-choice question that
 they were all expected to know, namely the colors of the
 American flag. The demographic questions queried age,
 gender, race, ethnicity, and political ideology, a scale from

528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576

Table 1. Type III Analysis of Variance table describing change in sentencing scores (in years) and satisfaction ratings as a function of evidence of perpetrator understanding, suffering, and crime seriousness

Measure	df	Estimate (SE)	F	p	Random effect (SD)
Intercept					
Sentencing Δ	1, 509.121	-0.397 (.106)	11.156	< .001	0.862
Satisfaction	1, 508.225	-1.771 (.129)	267.778	< .001	2.041
Understanding					
Sentencing Δ	1, 913.473	0.826 (.154)	181.981	< .001	
Satisfaction	1, 895.381	-0.394 (.186)	52.395	< .001	
Suffering					
Sentencing Δ	1, 908.750	0.052 (.152)	24.586	< .001	
Satisfaction	1, 866.854	0.089 (.183)	5.947	.015	
Crime Seriousness					
Sentencing Δ	1, 508.736	-0.016 (.131)	0.127	.722	
Satisfaction	1, 506.219	2.390 (.212)	381.919	< .001	
Understanding \times Suffering					
Sentencing Δ	1, 913.665	0.330 (.217)	8.046	.005	
Satisfaction	1, 901.012	-0.238 (.261)	0.809	.369	
Understanding \times Crime Seriousness					
Sentencing Δ	1, 908.185	-0.238 (.189)	2.074	.150	
Satisfaction	1, 864.515	-0.588 (.306)	6.454	.011	
Suffering \times Crime Seriousness					
Sentencing Δ	1, 913.348	0.176 (.187)	2.823	.093	
Satisfaction	1, 893.680	-0.506 (.305)	4.580	.033	
Understanding \times Suffering \times Crime Seriousness					
Sentencing Δ	1, 906.429	0.093 (.265)	0.124	.725	
Satisfaction	1, 858.679	0.084 (.429)	0.039	.844	

577 very liberal (-3) to very conservative (+3). See preregistration
578 protocol for the exact text of all items.

579 Results

580 Primary Hypothesis Tests

581 H1: Did Testimony That the Perpetrator Came to 582 Understand the Wrongfulness of His Actions 583 (a) Decrease Prison Sentence Recommendations 584 and (b) Increase Satisfaction With the Prospect 585 of Parole Punishment?

586 This hypothesis was fully supported. According to a linear
587 mixed model, a main effect of understanding was observed
588 such that sentence lengths decreased relative to baseline
589 when the perpetrator came to understand that what he
590 did was wrong compared to when he did not show such
591 understanding. To estimate the size of this main effect,
592 we computed a Pseudo- R^2 as recommended by Snijders
593 and Bosker (2011) by comparing the residual variance of
594 an empty model (i.e., excluding any predictors) with a
595 model containing only the understanding main effect (using
596 restricted Maximum Likelihood estimation). This effect was

597 large, explaining over 14% of the variance in sentencing
598 change scores ($R^2 = .146$). In addition, participants were
599 more satisfied with the prospect of early parole when the
600 perpetrator understood the wrongfulness of his actions than
601 when he did not. This effect was small to medium in size
602 ($R^2 = .025$) (see Table 1 and Figures 1, 2, 3, and 4; also,
603 Appendix B, available at <https://dx.doi.org/10.23668/psycharchives.5005>).

604 This pattern successfully replicated using the between-
605 subjects post-manipulation sentencing score, $F(1, 767.870)$
606 = 32.971, $p < .001$. Participants recommended shorter sen-
607 tence lengths when the perpetrator came to understand
608 that what he did was wrong, $M = 4.671$, $SD = 2.376$, 95%
609 CI [4.464, 4.878], compared to when he did not show such
610 understanding, $M = 5.408$, $SD = 2.408$, 95% CI [5.200,
611 5.616], $p < .001$.

613 H2: Did Testimony That the Perpetrator Suffered 614 From the Punishment (a) Decrease Prison Sentence 615 Recommendations and (b) Increase Satisfaction With 616 the Prospect of Parole Punishment?

617 Using the same linear mixed model, this hypothesis was
618 also fully supported. First, sentence lengths decreased rela-
619 tive to baseline when the perpetrator putatively suffered as

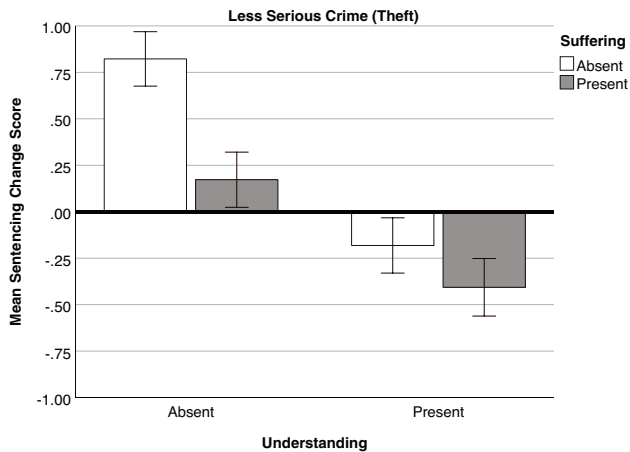


Figure 1. Estimated marginal mean values for sentencing change score (years) as a function of evidence of perpetrator understanding and suffering for the less serious crime (data are centered on the baseline punishment mean, $M = 2.505$ years, $SD = 1.746$).

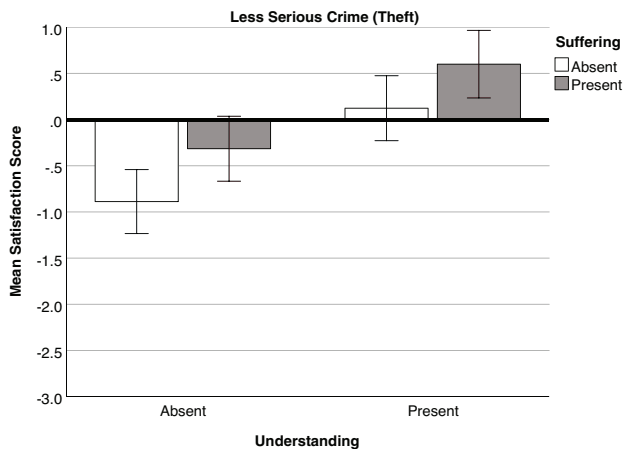


Figure 2. Estimated marginal mean values for satisfaction ratings as a function of evidence of perpetrator understanding and suffering for the less serious crime.

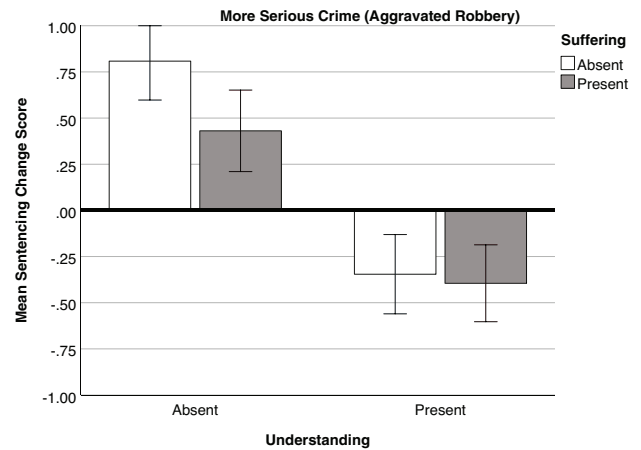


Figure 3. Estimated marginal mean values for sentencing change score (years) as a function of evidence of perpetrator understanding and suffering for the more serious crime (data are centered on the baseline punishment mean, $M = 7.584$ years, $SD = 2.743$).

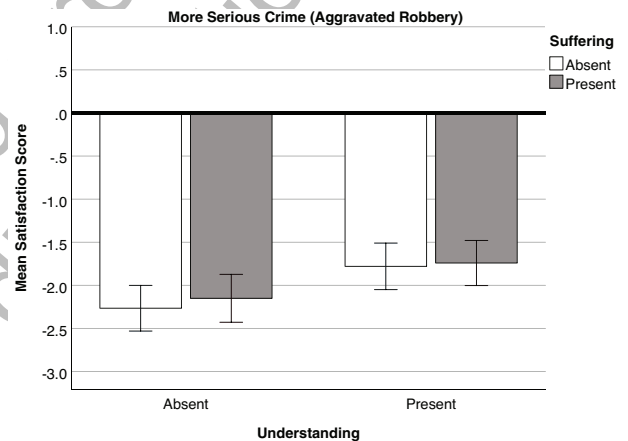


Figure 4. Estimated marginal mean values for satisfaction ratings as a function of evidence of perpetrator understanding and suffering for the high-seriousness crime.

620 a result of the punishment compared to when he did not
 621 suffer. The Pseudo- R^2 score (Snijders & Bosker, 2011)
 622 suggests a small to medium-size effect of suffering ($R^2 =$
 623 $.020$). Second, participants were more satisfied with the
 624 prospect of early parole when he suffered than when he
 625 did not. This effect was small ($R^2 = .004$) (see Table 1
 626 and Figures 1, 2, 3, and 4; also, Appendix B, available at
 627 <https://dx.doi.org/10.23668/psycharchives.5005>).

628 This pattern successfully replicated using the between-
 629 subjects post-manipulation sentencing score, $F(1,$
 630 $740.100) = 5.475$, $p = .020$. Participants recommended
 631 more lenient sentences when the perpetrator suffered as
 632 a result of the punishment, $M = 4.891$, $SD = 2.382$, 95%
 633 CI [4.684, 5.099], compared to when he did not suffer,
 634 $M = 5.188$, $SD = 2.401$, 95% CI [4.983, 5.393], $p = .020$.

H3: Did the Combined Presence of Suffering and Understanding Uniquely (a) Decrease Prison Sentence Recommendations and (b) Increase Satisfaction With the Prospect of Parole Punishment?

635 To test this hypothesis, we examined potential interactions
 636 between understanding and suffering using the same linear
 637 mixed model described above. The predicted synergistic
 638 interaction effect was not found. Participants were no more
 639 likely to reduce their sentences when understanding and
 640 suffering were both present compared to when just one of
 641 these was present. However, understanding and suffering
 642 interacted in another way, such that participants were espe-
 643 cially punitive when both understanding and suffering were
 644 absent—a synergistic interaction of the reverse direction
 645 than that predicted. Furthermore, the aggravating effect
 646
 647
 648
 649

of suffering only held when the perpetrator did not understand the wrongfulness of his actions. When he did exhibit understanding, the presence of suffering did not evoke changes in sentences. To estimate the size of this interaction effect, we compared the residual variance of a model that contained only the main effects of understanding and suffering with a model that also contained the interaction between the two. This effect was small ($R^2 = .007$). Regarding satisfaction ratings, no interactions between understanding and suffering were observed (see Table 1 and Figures 1, 2, 3, and 4; also, Appendix B, available at <https://dx.doi.org/10.23668/psycharchives.5005>). Moreover, post-manipulation sentencing scores failed to reproduce the understanding by suffering interaction observed using the sentencing change score, $F(1, 773.924) = 1.326, p = .250$.

To evaluate the null interaction observed for the satisfaction measure, we tested our model's assumption of normality. Since the satisfaction measure was not normally distributed ($D = .212, p < .001$), a log10 transformation was applied to approximate a normal distribution. However, the interaction between understanding and suffering remained non-significant while using the transformed measure, $F(1, 928.862) = 0.798, p = .372$.

Finally, we retested the interaction between understanding and suffering on satisfaction ratings only among participants who rated the perpetrator's suffering and understanding greater than zero on the manipulation check scales. However, an interactive effect did not emerge, $F(1, 116.706) = 1.634, p = .204$.

Exploratory Analyses

Did Crime Seriousness Moderate the Effects of Understanding and Suffering on Punishment Goal Fulfillment?

Using the same linear mixed model described above, we inspected interactive effects between crimes' seriousness and the other factors. Crime seriousness did not moderate the effects of understanding, suffering, or their interaction on sentencing change scores. However, satisfaction ratings showed evidence of moderation. First, crime seriousness exerted the main effect on satisfaction wherein participants were less dissatisfied by the prospect of early parole for the less serious crime (theft) than, the more serious crime (aggravated robbery). Second, the effect of understanding on satisfaction was significantly greater for the less serious crime than the more serious crime. Specifically, while the presence of perpetrator understanding significantly increased satisfaction with the prospect of early parole across both crimes, this increase was greater for the theft than the robbery. However, the simple effect of understanding on satisfaction for the robbery was still significant. Third, the effect of suffering on satisfaction was also

moderated by crime seriousness. In this case, evidence that the perpetrator suffered only increased satisfaction with the prospect of early parole for the theft and not for the robbery. That is, the simple effect of suffering on satisfaction for the theft was significant, but the equivalent test for the robbery was non-significant (see Table 1 and Figures 2 and 4; also, Appendix B, available at <https://dx.doi.org/10.23668/psycharchives.5005>). On the whole, the seriousness of the crime had a limited moderating effect, curtailing the effects of understanding and suffering on punishment goal fulfillment at higher levels (i.e., robbery) but only for the satisfaction ratings, and not for the sentencing scores.

Did Party Perspective (Personal Victim of Crime vs. Impersonal) Moderate the Effects of Understanding and Suffering on Punishment Goal Fulfillment?

To answer this question, we constructed a MANOVA for each of the low and high seriousness crimes, using understanding, suffering, and party perspective as independent factors and sentencing change scores and parole satisfaction scores as the dependent measures. No evidence of moderation by party perspective was found.

For the less serious crime, the multivariate main effect of party perspective on punishment goal fulfillment was not significant, $F(2, 505) = 2.730, p = .066, f^2 = .011$. Moreover, party perspective did not moderate the effect of understanding, $p = .957$, suffering, $p = .405$, or their interaction, $p = .312$.

For the more serious crime, a significant main effect of party perspective on punishment goal fulfillment was observed, $F(2, 505) = 4.920, p = .008, f^2 = .019$. This effect was limited to the satisfaction measure, $F(1, 506) = 9.710, p = .002, f^2 = .019$, wherein participants showed less satisfaction with the prospect of early parole when participants were portrayed as the victim ($M = -2.194, SD = 1.535, 95\% CI [-2.380, -2.007]$) than when they were not ($M = -1.771, SD = 1.543, 95\% CI [-1.961, -1.581]$). However, this pattern was not replicated using the sentencing change score, $p = .436$. Likewise, party perspective did not interact with understanding, $p = .177$, suffering, $p = .473$, or their interaction, $p = .340$. Thus, the effects of understanding and suffering on punishment goal fulfillment do not appear to depend on the punisher's perspective.

Individual Differences in Responsiveness to Testimony on Perpetrator Understanding and Suffering

Our baseline sentencing measure afforded a unique opportunity to examine what proportion of participants did and did not change their sentencing recommendation following the understanding and suffering manipulations. To our surprise, the majority of participants chose to persist in their original punishment judgment (67.32% for the less serious

crime and 79.77% for the more serious crime). This uniformity of sentencing is surprising for two reasons. First, it means that the substantial restriction in within-subject variability was not sufficient to negate the predicted effects, suggesting that these effects were quite robust. Second, it might suggest that perpetrator suffering and understanding, at least as presented in our manipulations, are not sufficient to satisfy these individuals' punishment goals.

To address this latter possibility, we examined whether the participants with uniform sentencing recommendations might still be more satisfied with the prospect of early parole when the perpetrator demonstrated evidence of understanding and/or suffering. If uniform punishers truly do not value information about perpetrator understanding or suffering, then this information should not affect their satisfaction with early parole. Indeed, in the case of perpetrator suffering, that information did not influence satisfaction among the uniform punishers, according to a two-way ANOVA for the less serious crime, $F(1, 342) = 2.304, p = .130$, or the more serious crime, $F(1, 406) = 2.294, p = .131$. However, evidence of perpetrator understanding did increase satisfaction with early parole among these uniform punishers, *both* for the less serious crime, $F(1, 342) = 9.072, p = .003, f^2 = .027$, and the more serious crime, $F(1, 406) = 4.047, p = .045, f^2 = .010$. These uniform punishers were less dissatisfied with the prospect of early parole when the perpetrator of the less serious crime displayed understanding, $M = 0.056, SD = 2.150, 95\% CI [-0.250, 0.362]$, than when he did not, $M = -0.648, SD = 2.186, 95\% CI [-0.991, -0.305]$. Similarly, they were less dissatisfied with the prospect of early parole when the perpetrator of the more serious crime showed evidence of understanding, $M = -1.978, SD = -1.427, 95\% CI [-2.166, -1.791]$, than when he did not, $M = -2.261, SD = 1.416, 95\% CI [-2.464, -2.058]$. Thus, we find evidence of punishment goal satisfaction when the perpetrator demonstrates understanding even among those who chose not to alter their sentencing recommendations (for supplemental analyses, see Appendix A, available at <https://dx.doi.org/10.23668/psycharchives.5005>).

Discussion

The purpose of this study was to examine two proximate drivers of punishment: evidence that perpetrators understood that what they did was wrong and felt remorse and evidence that they suffered as a result of the punishment. We hypothesized that credible evidence of understanding and suffering would independently and jointly reduce the length of recommended prison sentences and increase reported satisfaction with a relatively lenient sentence – both measures of punishment goal fulfillment.

Our overall pattern of results was generally consistent with these predictions, though with some evidence that understanding played a more substantial role than suffering in participants' punitive responses. These results indicated that, across levels of crime seriousness (theft vs. aggravated robbery) and party perspectives (personal vs. impersonal), sentencing recommendations were substantially lower when participants were given testimony that the perpetrator understood that what he did was wrong and felt remorse than when given testimony that he lacked such understanding and remorse. This effect held when we compared post-manipulation sentence recommendations directly and when we accounted for each participant's baseline punishment recommendation (before exposure to the manipulations). The results also revealed ~~increased satisfaction~~ with a relatively lenient punishment option (parole after 10 weeks in jail) when understanding was present versus absent. This difference in satisfaction emerged even among participants who did not reduce their own sentencing recommendation. This pattern of results conceptually replicates the "understanding effect" that previous research has found (Funk et al., 2014; Gollwitzer & Denzler, 2009; Gollwitzer et al., 2011; Molnar et al., 2020).

A similar pattern of results was found in response to perpetrator suffering, consistent with our predictions and, also, with previous findings (Eder et al., 2020), but the size of this effect was smaller than the one for understanding. Contrary to our predictions, we did not find that the combined presence of understanding and suffering mitigated punishment more strongly than just one or the other. However, we did find a small interaction effect in the "reverse" direction, namely that punishments were most severe when both suffering and understanding were absent compared to all other combinations of these variables. However, the aggravating effect of lack of suffering only held when the perpetrator did *not* understand the wrongfulness of his actions. These aggravating effects suggest that participants may have expected the perpetrator to have already understood and suffered by default, so evidence affirming this expectation would be less powerful than evidence refuting it.

It is noteworthy that the effects of understanding and suffering on satisfaction with a more lenient sentence were strongest for the less serious crime. This pattern suggests a possible floor effect whereby the more serious, violent crime evoked so much dissatisfaction with the prospect of early parole that the relatively minimal information provided to participants about the perpetrator's understanding or suffering during his 10-weeks in jail was insufficient to impact their sentencing judgment or their satisfaction with such a short time of incarceration. Perhaps for these more serious crimes, people demand a stronger degree of understanding or suffering, or perhaps they require more evidence that understanding or suffering has occurred. Adjusting the

strength of the manipulations as well as the extent of the more lenient punishment alternative (i.e., so it is not quite so lenient) could help test these possible explanations.

Taken together, both understanding and suffering predictably influenced sentencing judgments irrespective of the seriousness of the crime or the party perspective. But while understanding mitigated punishment, regardless of how much the offender suffered, the effect of suffering was conditional on the perpetrator's level of understanding, implicating a more direct and more stringent test of the "understanding hypothesis" than previous research has undertaken (e.g., Funk et al., 2014; Gollwitzer et al., 2011). To the extent that responsiveness to these factors reflects on punishment goal fulfillment, this pattern suggests that getting perpetrators to understand the wrongfulness of their actions and feel remorse might serve as a primary goal of punishment, whereas making the perpetrator suffer serves a secondary goal which becomes activated when the more primary goal of understanding has failed. It might even be that some people see suffering as a necessary means to induce remorse precisely when they see no evidence of understanding. In any case, this overall pattern of results suggests that ordinary punishment judgments may be shaped by motivations to induce both understanding and suffering.

Despite finding support for our hypotheses about the effects of understanding and suffering, we discovered that the majority of participants preferred not to change their original sentencing recommendations. One explanation for this high proportion of uniform sentences is the operation of a demand characteristic whereby the repeated sentencing measure roused participants' suspicions. But the fact that our hypothesis was still supported in the between-group (post-manipulation) analysis weakens this interpretation. It could also be that people held to their initial judgment because of belief perseverance, the tendency to maintain a belief after obtaining new evidence which controverts that belief (Anderson et al., 1980). But this explanation is complicated by the fact that these participants, like those who changed their sentences, were also found to be less dissatisfied by the prospect of early parole when the perpetrator exhibited understanding. This inconsistency suggests that increased satisfaction is not necessarily sufficient to motivate a change in overt punishment behavior, suggesting that there must be other motivational states that contribute to punishment judgments in these individuals that were not captured by the present study.

Limitations and Future Directions

This study included some known limitations that should be considered in future research. First, it utilized a hypothetical vignette method, which permitted us to experimentally

probe intuitions about criminal behavior. Previous studies have also found similar effects using behavioral games to study non-criminal social violations, but those studies employed a personal perspective only (e.g., Gollwitzer & Denzler, 2009; Molnar et al., 2020). Thus, there remains a demand to develop new, ecologically realistic ways to simulate personal and impersonal judgments of criminal behavior, such as mock jury studies.

This study tested two crime types varying in seriousness. Importantly, this pattern of results may not necessarily generalize to other crime types. But this points to a larger issue in this body of literature, which is the tendency to treat crime seriousness as a unidimensional construct. Manipulating crime type can result in changes to any number of important psychological cues, including attributions of criminal intent, damages, dangerousness, and more. Future studies of punishment motivations should consider ways to systematically control these separable aspects of a crime.

Our operational definitions of suffering and understanding contained important tradeoffs. For instance, our definition of suffering included second-hand evidence of a medical condition that was aggravated by incarceration. This suffering is, at best, a side effect rather than an intended effect of imprisonment. This could help to explain why many participants did not change their sentencing recommendations and why participants in both suffering conditions tended to disagree with the statement that the perpetrator genuinely suffered as a result of the 10 weeks in jail. We would expect that a more intentional effect of punishment on suffering would exhibit an even greater influence on punishment goal satisfaction. This expectation is consistent with previous research, which has found that seeing a perpetrator suffer (seemingly from fate) only increases punishment goal fulfillment if the suffering is interpreted by the perpetrator as punishment (Gollwitzer et al., 2011). However, efforts to increase perceived suffering as an intended effect of incarceration would likely pose problems for a study's ecological validity, since inflicting suffering, according to many participants and legal scholars, should not be an explicit goal of modern penal institutions. Future research, perhaps with a focus on vigilante justice, should anticipate and attempt to balance such tradeoffs.

Concluding Remarks

Our findings both replicate and qualify the "intuitive retributivism" hypothesis, according to which people, when thinking about appropriate punishment for another person's wrongdoing, are concerned with re-balancing the scales of justice. Our results are also consistent with other empirical research showing that cues of understanding, remorse and apology increase punishment satisfaction (Bauer & Poama, 2020; Funk et al., 2014; Gollwitzer & Denzler,

2009; Gollwitzer et al., 2011; Molnar et al., 2020). These findings suggest the psychological plausibility of communicative theories of punishment, which argue that the goals of punishment are to send a message to perpetrators that they have violated a community norm or law and that they must respond appropriately with regret, apology, and ideally offers of reparation and attempts at rehabilitation. Such theories hence predict that these aims will be achieved with evidence that the perpetrator understands those messages and responds accordingly and that punishers will respond with diminished punishment and ideally forgiveness and reintegration of the perpetrator into the community (see, e.g., Berman, 2010; Cushman, 2015; Cushman et al. [author: please update], in press; Duff, 2001; Funk et al., 2014; McGeer & Funk, 2017; Nahmias & Aharoni, 2017). Thus, people are “intuitive retributivists” in the sense that they aim at closing the injustice gap that the wrongdoing has opened. And communicating disapproval and effecting a change in the offender does close this injustice gap, even without making the offender suffer. Second, people are also “intuitive retributivists” in the sense that seeing or making the offender suffer from the punishment is also **rebalancing** the scales of justice, but this specific concern is only of secondary importance: it begins to play a role when the primary goal – making the offender understand that what he did was wrong – has failed. With this particular finding, our study also lends support to research on the role of perpetrator suffering in punishment judgments (Eder et al., 2020; Sinaceur et al., 2015). However, the effects of suffering were notably smaller than the effects of understanding, which also qualifies the notion that people’s “sense of justice derives . . . from reflection upon comparative suffering” (Frijda, 1994; pp. 274–275).

In addition to replicating previous research examining the psychological mechanisms of punishment, this study makes several original contributions to the punishment literature. First, this is the first study of its kind to compare suffering with understanding for multiple crime types and from both personal and impersonal perspectives. This allowed us to evaluate the relative contributions of suffering and understanding in a variety of contexts. Second, we employed a large sample, approximating attributes of the US population along basic demographic lines. Third, our design also included both between- and within-subjects measurements of punishment goal fulfillment, including both behavioral and attitudinal measures, which demonstrated good internal consistency.

The results of this study may also have some relevance for legal practitioners. If it is true, for instance, that people are more sympathetic toward defendants who display evidence of suffering but only for less serious crimes, defense lawyers could factor this information into their defense strategy. On the other hand, if the evidence suggests that

perpetrator suffering is not a necessary component of punishment goal fulfillment or that genuine understanding may be sufficient, then lawmakers seeking to represent the values of ordinary people might find justification in sentencing policies that emphasize rehabilitation and restitution over retribution and deterrence. To achieve these aims, legal systems might need to develop more systematic processes to allow perpetrators to demonstrate remorse and apology and make amends, as restorative justice reforms have aimed to do. While some have presented these reforms as *alternatives* to punishment, communicative theories can adopt them as ways to satisfy people’s most essential aims in punishing those who violate the community’s norms – namely, demanding that they come to understand the harms they caused, to feel remorse for them, and to try to correct those harms.

References

- Aharoni, E., & Fridlund, A. J. (2012). Punishment without reason: Isolating retribution in lay punishment of criminal offenders. *Psychology, Public Policy, and Law*, 18(4), 599–625. <https://doi.org/10.1037/a0025821>
- Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, 39(6), 1037–1049. <https://doi.org/10.1037/h0077720>
- Bauer, P. C., & Poama, A. (2020). Does suffering suffice? An experimental assessment of desert retributivism. *PLoS One*, 15(4), Article e0230304. <https://doi.org/10.1371/journal.pone.0230304>
- Berman, M. N. (2010). *Two kinds of retributivism* (Public Law Research Paper No. 171). University of Texas Law.
- Carlsmith, K. M. (2006). The roles of retribution and utility in determining punishment. *Journal of Experimental Social Psychology*, 42(4), 437–451. <https://doi.org/10.1016/j.jesp.2005.06.007>
- Carlsmith, K. M., & Darley, J. M. (2008). Psychological aspects of retributive justice. *Advances in Experimental Social Psychology*, 40, 193–236. [https://doi.org/10.1016/S0065-2601\(07\)00004-4](https://doi.org/10.1016/S0065-2601(07)00004-4)
- 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., Monahan, J., & Evans, A. (2007). The function of punishment in the “civil” commitment of sexually violent predators. *Behavioral Sciences & the Law*, 25(4), 437–448. <https://doi.org/10.1002/bsl.761>
- 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>
- Crombag, H., Rassin, E., & Horselenberg, R. (2003). On vengeance. *Psychology, Crime & Law*, 9, 333–344. <https://doi.org/10.1080/1068316031000068647>
- Cushman, F. (2015). Punishment in humans: From intuitions to institutions. *Philosophy Compass*, 10(2), 117–133. <https://doi.org/10.1111/phc3.12192>
- Cushman, F., Sarin, A., & Ho, M. (in press). [author: please update] Punishment as communication. In J. Doris & M. Vargas (Eds.), *The Oxford handbook of moral psychology*. Oxford University Press. <https://psyarxiv.com/wf3tz>

- 1069 Darley, J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and just deserts as motives for punishment. *Law and Human Behavior*, 24(6), 659–683. <https://doi.org/10.1023/A:1005552203727>
- 1070 Duff, R. A. (2001). *Punishment, communication, and community*. Oxford University Press.
- 1071 Duff, R. A. (2002). [Author: Please add publisher] Punishing the young. In I. Weijers & R. A. Duff (Eds.), *Punishing Juveniles: Principle and Critique* (pp. 115–134).
- 1072 Eder, A., Mitschke, V., & Gollwitzer, M. (2020). What stops revenge taking? Effects of observed emotional reactions on revenge seeking. *Aggressive Behavior*, 46(4), 305–316. <https://doi.org/10.1002/ab.21890>
- 1073 Ellsworth, P. C., & Ross, L. (1983). Public opinion and capital punishment: A close examination of the views of abolitionists and retentionists. *Crime & Delinquency*, 29, 116–169. <https://doi.org/10.1177/001112878302900105>
- 1074 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](https://doi.org/10.1016/S1090-5138(04)00005-4)
- 1075 Frijda, N. H. (1994). The Lex Talionis: On vengeance. In S. H. M. van Goozen, N. E. van der Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 263–289). Erlbaum.
- 1076 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>
- 1077 Gerber, M. M., & Jackson, J. (2013). Retribution as Revenge and Retribution as Just Deserts. *Social Justice Research*, 26(1), 61–80. <https://doi.org/10.1007/s11211-012-0174-7>
- 1078 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>
- 1079 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>
- 1080 Gromet, D. M., & Darley, J. M. (2009). Punishment and beyond: Achieving justice through the satisfaction of multiple goals. *Law & Society Review*, 43(1), 1–38. <https://doi.org/10.1111/j.1540-5893.2009.00365.x>
- 1081 Kleinke, C. L., Wallis, R., & Stalder, K. (1992). Evaluation of a rapist as a function of expressed intent and remorse. *The Journal of Social Psychology*, 132(4), 525–537. <https://doi.org/10.1080/00224545.1992.9924732>
- 1082 Marshall, J., Yudkin, D. A., & Crockett, M. J. (2020). [author: please update] Children punish third parties to satisfy both consequentialist and retributive motives. *Nature Human Behaviour*, 1–8. <https://doi.org/10.1038/s41562-020-00975-9>
- 1083 McFatter, R. M. (1982). Purposes of punishment: Effects of utilities of criminal sanctions on perceived appropriateness. *Journal of Applied Psychology*, 67(3), 255–267. <https://doi.org/10.1037/0021-9010.67.3.255>
- 1084 McGeer, V., & Funk, F. (2017). Are “optimistic” theories of criminal justice psychologically feasible? The probative case of civic republicanism. *Criminal Law and Philosophy*, 11, 523–544. <https://doi.org/10.1007/s11572-015-9381-2>
- 1085 Miller, D. T. (2001). Disrespect and the experience of injustice. *Annual Review of Psychology*, 52, 527–553. <https://doi.org/10.1146/annurev.psych.52.1.527>
- 1086 Molnar, A., Chaudhry, S., & Loewenstein, G. F. (2020). *It's not about the money. It's about sending a message!: Unpacking the components of revenge*. <https://ssrn.com/abstract=3524910>
- 1087 Moore, M. S. (1997). *Placing blame: A general theory of criminal law*. Oxford University Press.
- 1088 Nadelhoffer, T., Heshmati, S., Kaplan, D., & Nichols, S. (2013). Folk retributivism and the communication confound. *Economics and Philosophy*, 29(2), 235–261. <https://doi.org/10.1017/S0266267113000217>
- 1089 Nahmias, E., & Aharoni, E. (2017). Communicative theories of punishment and the impact of apology. In C. W. Surprenant (Ed.), *Rethinking punishment in an era of mass incarceration* (pp. 144–161). Routledge.
- 1090 Orth, U. (2004). Does perpetrator punishment satisfy victims' feelings of revenge? ~~Aggressive Behavior~~. *Aggressive Behavior* [Author: Please check & approve the source title], 30(1), 62–70. <https://doi.org/10.1002/ab.20003>
- 1091 Sinaceur, M., Kopelman, S., Vasiljevic, D., & Haag, C. (2015). Wee, and get more: When and why sadness expression is effective in negotiations. *Journal of Applied Psychology*, 100(6), 1847–1871. <https://doi.org/10.1037/a0038783>
- 1092 Smith, A. (2014). *Older adults and technology use. Adoption is increasing, but many seniors remain isolated from digital life*. Pew Research Center. <http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/>
- 1093 Snijders, T. A., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.
- 1094 Twardawski, M., & Hilbig, B. E. (2020). The motivational basis of third-party punishment in children. *PLoS One*, 15(11), Article e0241919. <https://doi.org/10.1371/journal.pone.0241919>
- 1095 Vidmar, N. (1974). Retributive and utilitarian motives and other correlates of Canadian attitudes toward the death penalty. *Canadian Psychologist/Psychologie Canadienne*, 15(4), 337–356. <https://doi.org/10.1037/h0081769>
- 1096 Vidmar, N., & Miller, D. T. (1980). [author: please check & approve vol and issue] Social psychological processes underlying attitudes toward legal punishment. *Law and Society Review*, 14(3), 565–602. <https://doi.org/10.2307/3053193>
- 1097 Vuk, M., Applegate, B. K., Ouellette, H. M., Bolin, R. M., & Aizpurua, E. (2020). The pragmatic public? The impact of practical concerns on support for punitive and rehabilitative prison policies. *American Journal of Criminal Justice*, 45(2), 273–292. <https://doi.org/10.1007/s12103-019-09507-2>
- 1098 Warr, M., Meier, R. F., & Erickson, M. L. (1983). Norms, theories of punishment, and publicly preferred penalties for crimes. *Sociological Quarterly*, 24(1), 75–91. <https://doi.org/10.1111/j.1533-8525.1983.tb02229.x>
- 1099 Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143(5), 2020–2045. <https://doi.org/10.1037/xge0000014>
- 1100 Zhou, Y., Jiao, P., & Zhang, Q. (2017). Second-party and third-party punishment in a public goods experiment. *Applied Economics Letters*, 24(1), 54–57. <https://doi.org/10.1080/13504851.2016.1161709>

History

Received April 3, 2020

Revision received January 22, 2021

Accepted April 26, 2021

Published online XX, 2022

Acknowledgments


All study methods were preregistered at PsychArchives. We thank Respondi for participant recruitment and data collection, PsychLab, a service of the ZPID – Leibniz Institute for Psychology for financial support, Robert Moulder and Sharlene Fernandes for analytic support, and the members of the Cooperation, Conflict, and Cognition Lab including Sharlene Fernandes, Corey H. Allen, Saskia Verkiel, and Angela Wingers for helpful comments on drafts of this manuscript.

Open Data

The complete pre-registration protocol and study materials are available at <http://dx.doi.org/10.23668/psycharchives.3160>.
Supplementary material (Appendices A and B) is available at <http://dx.doi.org/10.23668/psycharchives.5005>.

ORCID

Eyal Aharoni

 <https://orcid.org/0000-0003-3544-1247>

Eyal Aharoni

Department of Psychology
Georgia State University
PO Box 5010
Atlanta, GA 30302-5010
USA
eaharoni@gsu.edu

Uncorrected proof
- not for distribution