Supplementary materials

What Drives Second- and Third-Party Punishment? Conceptual Replications of the 'Intuitive Retributivism' Hypothesis

Mathias Twardawski[1], Mario Gollwitzer[1], Steffi Pohl[2], and Michael Bosnjak[3,4]

[1] Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

[2] Department of Psychology, FU Berlin, Berlin, Germany

[3] ZPID – Leibniz Institute for Psychology, Trier, Germany

[4] Department of Psychology, University of Trier, Germany

Technical details on the meta-analysis

The meta-analysis was carried out using either the correlation coefficient between the respective treatment and outcome variables (for the studies by Aharoni et al., this issue; De Cristofaro & Giacomantonio, this issue; Hechler & Kessler, this issue; Nockur et al., this issue; Molho et al., this issue) or differences between punishment motives measured directly or indirectly (for the studies by Fousiani & van Prooijen, this issue; Rehren & Zisman, this issue; Strauß & Bondü, this issue). All effect sizes were transformed to correlation coefficients. A random-effects model was fitted to the data. The amount of heterogeneity (i.e., τ^2), was estimated using the restricted maximum-likelihood estimator (Viechtbauer, 2005). In addition to the estimate of τ^2, the Q-test for heterogeneity (Cochran, 1954) and the I^2 statistic (Higgins & Thompson, 2002) are reported and a credibility/prediction interval for the true outcomes is also provided (Riley, Higgins, & Deeks, 2011). Studentized residuals and Cook's distances are used to examine whether studies may be outliers and/or influential in the context of the model (Viechtbauer & Cheung, 2010). Studies with a studentized residual larger than the $100 \times (1 - 0.05/(2 \times k))$th percentile of a standard normal distribution are considered potential outliers (i.e., using a Bonferroni correction with two-sided $\alpha = 0.05$ for k studies included in the meta-analysis). Studies with a Cook's distance larger than the median plus six times the interquartile range of the Cook's distances are considered to be influential. The rank correlation test (Begg & Mazumdar, 1994) and the regression test (Sterne & Egger, 2005), using the standard error of the observed outcomes as predictor, are used to check for funnel plot asymmetry. The analysis was carried out using R (version 4.0.3) and the metafor package (version 2.4.0) (Viechtbauer, 2010).

A total of $k = 9$ effect sizes from 8 studies were included in the analysis. One study was not included in the meta-analysis, as no effect size could be extracted that directly tested the evidence of the "intuitive retributivism" hypothesis (rather, this study was designed to falsify the "intuitive retributivism" hypothesis and, thus, did not provide effects on the presence of intuitive retributivism; Funk & Mischkowski, this issue). The observed correlation coefficients ranged from 0.0100 to 0.6100, with all estimates being positive. The estimated average correlation coefficient based on the random-effects model was $\hat{\mu} = 0.2644$ (95% CI: 0.1271 to 0.4017). Therefore, the average outcome differed significantly from zero ($z = 3.7748$, $p = 0.0002$).

An examination of the studentized residuals revealed that none of the studies had a value larger than ± 2.7729 and hence there was no indication of outliers in the context of this model. According to the Cook's distances, none of the studies could be considered to be overly influential.

A funnel plot of the estimates is shown in Figure S1. The regression test indicated funnel plot asymmetry ($p = 0.0484$) but not the rank correlation test ($p = 0.6122$).

According to the Q-test, the true outcomes appear to be heterogeneous ($Q(8) = 199.6775$, $p < 0.0001$, $\hat{\tau}^2 = 0.0414$, $I^2 = 95.5679\%$). A 95% credibility/prediction interval for the true outcomes is given by -0.1574 to 0.6861. Hence, although the average outcome is estimated to be positive, in some studies the true outcome may in fact be negative.

Figure S1.

Funnel plot

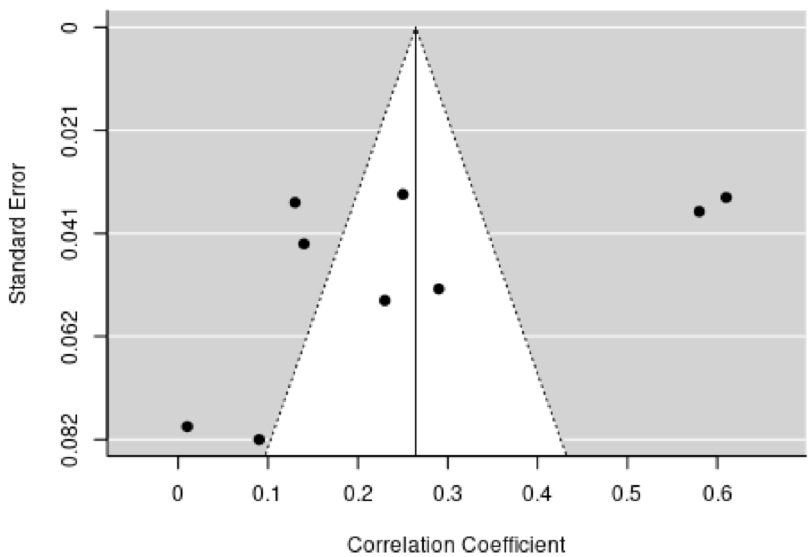

Table S1.

Effect sizes extracted from the articles that were used for the meta-analysis

Authors	**Description of the effects used for the meta-analysis**	***N***	**Effect sizes as reported in the manuscript**
Molho et al.	Main effect of offense severity (between-subjects factor) on punishment severity	308	$F(1, 303) = 26.90$, $p < .001$, $\eta^2 = .08$
De Cristofaro & Giacomantonio (1)	Correlation between intentionality of misbehavior with punishment severity	789	$r = .25$
De Cristofaro & Giacomantonio (2)	Correlation between intentionality of misbehavior with punishment inevitability	789	$r = .13$
Nockur et al.	Main effect of intentionality (within-subjects factor) on (hidden) punishment severity	146	$d = 0.18$
Hechler & Kessler	Main effect of intentionality (within-subjects factor) on punishment severity	302	$F(1,300) = 16.58$ $p < .001$, $\eta^2_p = .05$
Aharoni et al.	Main effect of offender suffering (between-subjects factor) on punishment severity	514	*pseudo* $R^2 = .020$
Rehren & Zisman	Testing the retributivism rank-preference score in the control condition against the expected value in a baseline model (= 5)	342	Retributivism rank-preference score ($M = 9.18$; $SD = 2.69$): $d = 1.55$
Fousiani & van Prooijen	Contrast between self-reported retributive vs. utilitarian and restorative motives	326	$F(1, 317) = 164.70$, $p < .001$, $\eta^2 = .34$
Strauß & Bondü	Main effect of the motive (retribution vs. utilitarianism) underlying a punishment reaction (within-subjects factor) on support of this reaction	170	$F(1, 157) = 0.02$, $p = .889$, $\eta^2 p < .001$