A R T I C L E

the british
psychological society
promoting excellence in psychology

When empathy leads to aggression: The effects of empathy on punitive attitudes towards aggressors

Célia F. Camara ⬤ | Alejandra Sel | Paul H. P. Hanel

Department of Psychology, University of Essex, Colchester, UK

Correspondence
Célia F. Camara, Department of Psychology, University of Essex, Colchester, UK.
Email: c.camaraperezvera@essex.ac.uk;
camaracelia@outlook.com

Abstract

When witnessing aggression, individuals often empathize more with victims than with aggressors, which may bias their perceptions and interpretations of the transgressions. However, the mechanisms underlying these biases remain poorly understood. Through two experiments, we investigated whether people's decisions to condemn aggressors are influenced by their predisposition to sympathize with the victim and explored how negative sentiments towards the aggressor may influence these decisions. Further, we tested the moderating role of callous-unemotional traits, hypothesizing that moral judgements and decisions to punish may differ among individuals who are less emotionally responsive, as they are less likely to sympathize with victims. Our findings revealed that greater empathy for victims intensified punitive attitudes towards aggressors, primarily mediated by participants' negative evaluations of the aggressor. Notably, such empathic inclinations were less prevalent among individuals with higher levels of callous-unemotional traits, as reflected by their lower concern for victims and greater inclination towards harsh punishments. These results offer insights into how justice-related attitudes may be shaped and potentially biased by individual differences in emotional responsiveness.

K E Y W O R D S
aggression, bias, empathy, moral judgement, punishment

INTRODUCTION

> An examination of empathy affords a powerful site from which to rethink the fundamental commitments of social science broadly and the sociology of punishment specifically.
>
> (Brown, 2012, p. 383)

The exploration of human responses to perceived injustice reveals the complexities embedded within our collective moral fabric. For example, brutal aggressions towards innocent victims often elicit a shared sense of moral outrage that often leads to demands for severe punitive measures (Carlsmith & Darley, 2008). Such an inclination is rooted in the popular belief that punishment should mirror the severity of the suffering inflicted on victims (Bastian et al., 2011; Carlsmith et al., 2002), suggesting a bias in the extent to which individuals empathize with victims over perpetrators. In the present research, we investigate how such biases might be determined by the perceived intentionality of the transgression, while also exploring how judgements of the transgressor might further influence punitive attitudes. The goal is to gain insights into how individual's predispositions to selectively apply empathy towards victims over perpetrators might shape their reactions to perceived injustice.

The empathic bias in moral judgement

Empathy has been broadly defined by its affective and cognitive facets, comprising the capacity to share and understand the subjective experience of others in reference to oneself (Decety, 2010). However, empathy can also refer to the capacity to feel concern about the welfare of others, which plays a key role in guiding moral behaviour (Decety et al., 2012; Decety & Jackson, 2004). Theoretically, it has been proposed that empathic concern prompts individuals to engage in charitable and altruistic behaviours (e.g. Batson et al., 1983; Underwood & Moore, 1982), as they are more prone to consider the well-being of others. Empathic concern has also been proposed to guide moral decision-making, helping internalize right from wrong by considering the perspectives, feelings and needs of others (Kohlberg, 1981; Piaget, 1932). Nevertheless, under specific circumstances, directing empathy towards some individuals may paradoxically diminish our empathy towards others (Simas et al., 2020). For example, people tend to show stronger empathy and helping intentions for ingroup members, which has been correlated with less empathetic responses towards those perceived as outgroup members (e.g. Cikara & Fiske, 2013; Tarrant et al., 2009, 2012; Vanman, 2016). Notably, this bias also seems to contribute to prejudice and discrimination against outgroups (Lalonde & Silverman, 1994; Tyler & Blader, 2003; Vanman, 2016).

In the context of aggression, empathic biases can look like taking the victim's side over the perpetrator's. By adopting the victim's perspective, individuals might engage in self-referential cognitive processes, amplifying their sensitivity to the victim's suffering as well as a sense of injustice on the victim's behalf (Ames et al., 2008; Ruby & Decety, 2004). These reactions may in turn foster negative evaluations of the perpetrator, thereby influencing decisions to punish them (e.g. Lin et al., 2024). A recent study by Lu and McKeown (2018) provides some evidence for this. Using a Dictator Game task, the authors found that participants with higher levels of empathic concern were more predisposed to punish unfair outgroup members than they were to compensate ingroup members. Similarly, in an earlier study, participants' expressions of empathic anger on behalf of a victim of aggression significantly predicted their intentions to punish the perpetrator (Vitaglione & Barnett, 2003). These findings provide important insights as to how empathy, in some contexts, might indeed lead to aggression (Batson, 1997; see also Neuberg et al., 1997).

It is important to note, however, that the application of punishment as a penalty for an immoral action can be seen as a social response seeking to maintain societal harmony, and thus can be described as aggression with prosocial motives (Levy, 2022). At the same time, these prosocial motivations differ across individuals (Shichman & Weiss, 2022), meaning that the underlying nature

as to why someone decides to punish a transgression might be subject to their inherent biases and perceptions.

The psychology of punishment

Research in psychology shows that people's evaluation of transgressions is often shaped by the extent to which blame is ascribed to the transgressor (e.g. Feather, 1994; Feather & Dawson, 1998; Weiner, 1995). As such, the same act may be judged differently depending on the perceived intentions of the perpetrator, as demonstrated in studies using 'moral sensitivity' tasks (e.g. Baez et al., 2014, 2017; Decety et al., 2012; Santamaría-García et al., 2017; Young & Saxe, 2008). In these tasks, participants are typically presented with scenarios where a transgression occurs either accidentally or intentionally. Findings consistently reveal that participants are more likely to morally condemn intentional transgressions compared to accidental ones (Baez et al., 2014; Bastian et al., 2011). Conversely, judgements of aggression may be more lenient when aggressors are perceived as not responsible for their actions or when the aggression is justified for a 'greater good' (Mikula, 2003).

This aligns with the Attribution of Blame model described by Shaver and Shaver (1985), which posits that assessments of injustice hinge on the degree of responsibility attributed to the perpetrator for violating someone else's entitlements without adequate justification. For example, while intentional harm to innocent victims is condemned and penalized (Mikula, 2003), aggressions towards violent offenders—often enforced through extreme forms of punishment such as capital penalties or castration—tend to be more justified and endorsed by the public (e.g. Bastian et al., 2013; Viki et al., 2012). This justification hinges on the idea that justice is implemented when punishments are proportional to the harm inflicted, with harsher punishments often perceived as effective strategies to both prevent the offender from reoffending as well as to discourage others from committing similar crimes. However, it should not be ignored that punishment may also be an emotional response to perceived injustice, which may drive individuals to advocate for (and potentially engage in) more severe measures for retaliation.

According to the Retributive Justice model proposed by Carlsmith et al. (2002), negative emotions arising from perceived injustice towards victims can lead to dehumanizing perpetrators (see also Darley & Pittman, 2003). This process may occur when perpetrators are seen as violating moral principles or social norms, justifying perceptions of them as less human and undeserving of basic rights (Haslam et al., 2005). Such judgements gain special relevance in legal contexts, where the severity of punishment for the wrongdoer is often based on the perceived intentionality of the committed crime (Bastian et al., 2011). This can lead to the endorsement of harsher punitive measures and reduced advocacy for the reform and reintegration of violent offenders, particularly in cases where perpetrators are perceived as inherently immoral or beyond redemption (Bastian et al., 2013; Harris & Rice, 2006; Osofsky et al., 2005). For instance, Viki et al. (2012) found that dehumanizing sex offenders positively correlated with the public's desire to exclude them from society, reduced support for their rehabilitation and increased endorsement of castration as a form of punishment (see also Bastian et al., 2013). These propositions raise important questions about the extent to which empathic biases and negative character judgements might influence moral reasoning and justice-related decisions:

> **Hypothesis 1.** We hypothesize that greater empathy towards victims will be associated with harsher punitive attitudes, especially when the harm is perceived as intentional (Baez et al., 2014; Bastian et al., 2011). We propose that this relationship will be mediated by the level of concern for the perpetrator, and the extent to which the perpetrator is perceived as 'mean'. Specifically, we hypothesize that higher levels of empathy for victims will correlate with a decreased level of concern for perpetrators as well as an increased perception of the perpetrator's meanness, both of which are expected to be related to harsher punitive attitudes towards perpetrators (Bastian et al., 2013; Decety & Cowell, 2015; Viki et al., 2012).

Individual differences in perceptions of injustice

Research suggests that people's responses to perceived injustice may vary depending on their emotional involvement when making moral judgements (Brown, 2012; Decety & Cowell, 2015; Mikula, 2003). Therefore, an individual's capacity (or willingness) to emotionally engage with others' experiences is key when examining their drive to retaliate on the behalf of others. Compelling examples can be found in the context of psychopathy. People with psychopathic tendencies are typically less sensitive to perceived injustice (Decety & Yoder, 2016). For example, individuals with psychopathy often engage in calculated and goal-oriented aggression (e.g. using manipulation or deceit to achieve personal gain) even when fully aware that their actions cause harm to others (Gini et al., 2015). Similarly, studies have shown that psychopaths can correctly identify morally wrong actions but fail to show guilt or remorse for their wrongdoings (Koenigs et al., 2012). To put it simply, '(p)sychopaths know what is right or wrong, but simply don't care' (Cima et al., 2010, p. 66).

Such disregard for others has been attributed to the expression of callous-unemotional traits (Lockwood et al., 2013). Callous-unemotional traits denote a cluster of personality characteristics marked by a lack of empathy, remorse and guilt, alongside shallow affect and a diminished sensitivity to others' emotional experiences (Frick et al., 2003). Individuals exhibiting these traits—even at subclinical levels—tend to approach moral dilemmas with a more detached and utilitarian perspective (Koenigs et al., 2012). Callous-unemotional traits have also been linked to a greater tendency to justify or normalize aggressive behaviour, particularly in situations involving personal gain or social dominance (Barchia & Bussey, 2011; Gini et al., 2015). Such tendencies could shift moral judgements, potentially reducing feelings of guilt or remorse in the face of immoral actions and perpetuating harmful behaviours (Čehajić et al., 2009; Perren & Gutzwiller-Helfenfinger, 2012).

That is not to say, however, that all individuals exhibiting callous-unemotional traits are inherently immoral and aggressive (Campos et al., 2022). In fact, research on successful psychopathy indicates that some psychopathic traits may be advantageous in certain circumstances (Hall & Benning, 2006). For example, individuals with callous-unemotional traits may be less influenced by emotional appeals when making moral judgements (Fragkaki et al., 2016; Vasconcelos et al., 2021). This reduced susceptibility to emotional distress, in turn, can be beneficial in professions that require emotional detachment, such as law enforcement (Dutton, 2012). From this perspective, one could argue that callous-unemotional traits might reduce affective biases during punishment decision-making, although no previous study has investigated this. We hypothesize the following:

> **Hypothesis 2.** Extending *Hypothesis 1*'s prediction that greater concern for victims will correlate with harsher punitive judgements, we further hypothesize that individual differences in callous-unemotional traits will attenuate this effect. Specifically, we propose that the positive relationship between victim-focused concern and punishment severity will be significantly weaker at higher levels of callous-unemotional traits, given that individuals with these traits are less likely to feel concern for victims (e.g Decety & Yoder, 2016). Additionally, we explore whether these hypothesized effects might influence punitive attitudes as well.

Overview of studies

We conducted two studies to test our hypotheses. Study 1 required participants to evaluate scenarios depicting interpersonal harm, both intentional and accidental. Participants were asked to express their concerns for both the victim and the perpetrator, assess the perceived 'meanness' of the perpetrator, and determine appropriate levels of punishment. Study 2 replicated this task, adding a measure of participants' self-reported levels of callous-unemotional traits to explore how responses in the task could be influenced by the expression of these traits. Data and analysis code can be accessed at https://osf.io/

bfmnt/?view_only=180a8ea9aea140b69f97af6b4b238ffc. Data were analysed using IBM SPSS Statistics for Windows, version 29.0 and *R*, version 4.0.0 (R Core Team, 2020). Study design and analysis were not preregistered.

STUDY 1

Method

Participants

A power analysis indicated that 111 participants were required to detect a small-to-medium effect size ($f = 0.175$) in a 3×2-within-subjects design with 95% power. A total of 126 participants were initially recruited at the University of Essex (UK). However, 15 participants did not complete the task, and one was excluded for failing attention checks, resulting in a final sample of 110 volunteers. The sample primarily consisted of university students aged 18–33 years ($M = 21.69$, $SD = 2.77$). The group included 76 women, 34 men, and one participant identifying as non-binary/third gender. Ethnically, 61.3% identified as White, 21.6% as Asian/Asian British, 7.2% as Black, and 9.9% as belonging to multiple or other ethnic groups. Ethical approval was obtained from the first author's institution.

Experimental design and stimuli

After giving their informed consent, participants completed a *Moral sensitivity task* (adapted from Decety et al., 2012), designed to measure their responses to different types of transgressions (see Figure 1 for examples). These included the following: (a) scenarios where one person purposefully causes harm to another person (intentional condition); (b) scenarios where one person unintentionally causes harm to another person (accidental condition); and (c) scenarios involving intentional harm directed at an object
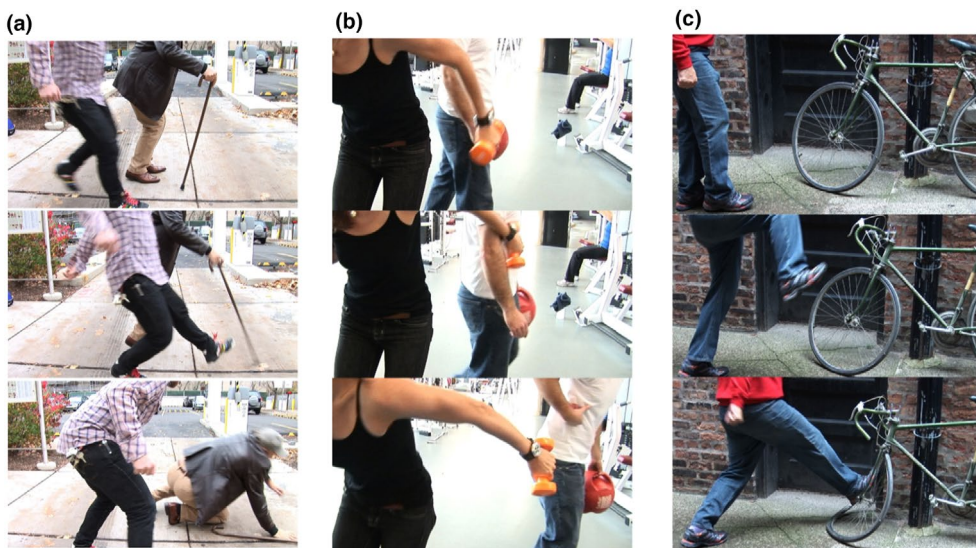
FIGURE 1 Examples of animated images in moral sensitivity task (Studies 1–2). Panel (a) illustrates a scenario depicting intentional harm to another person, representing the intentional condition. Panel (b) depicts a scenario portraying accidental harm to another person, representing the accidental condition. Lastly, panel (c) showcases a scenario demonstrating intentional harm to an object, serving as the control condition.

rather than a person (control condition). After each scenario, participants selected 'Yes' or 'No' to indicate whether the transgression was deliberate or unintentional ('was the action done on purpose?'). Subsequently, they used visual analogue scales ranging from 0 to 100 to rate their concern for both the victim ('how sorry do you feel for the injured person/damaged object?') and the perpetrator ('how sorry do you feel for the perpetrator?'), as well as their judgement of the perpetrator's meanness ('how mean was the perpetrator?') and appropriate punishment ('how harshly would you punish the perpetrator?').

The task comprised 2 initial practice trials, followed by 6 experimental trials (2 per condition, randomized across participants). The stimuli were created using 3 digital colour images, shown sequentially at durations of 500, 200 and 1000 ms to imply motion. Participants were blinded to the protagonists' faces to avoid facial bias, but the victim's emotional reaction and the perpetrator's intent remained inferable through body language. To reduce gender bias, the conditions included interactions across male-to-male, female-to-female, male-to-female and female-to-male pairings.

Analysis strategy

First, we assessed the success of our manipulation by evaluating participants' accuracy in distinguishing between intentional and accidental conditions, with successful manipulation defined as correctly identifying intentionality in more than 50% of the trials. Next, we tested our hypothesis that participants would express greater concern for victims than for perpetrators by conducting a 2 (victim vs. perpetrator) $\times$ 3 (control vs. intentional vs. accidental) repeated-measures ANOVA. Significant interactions were further explored using simple main effects analyses. Additionally, we examined differences in perceived perpetrator meanness and punishment ratings across conditions using one-way repeated-measures ANOVAs. For all analyses, follow-up pairwise comparisons were conducted using a Holm-Bonferroni correction, with an alpha level set at 0.05 and partial eta-squared (η_p^2) reported as the effect size. We also reported 95% confidence intervals (95% CI) where applicable.

To test our hypotheses about how perceived perpetrator meanness and concern for perpetrators influence the relationship between concern for victims and punishment, we used structural equation modelling (SEM) with the lavaan package in R (Rosseel, 2012). We initially introduced the variables measuring concern for perpetrators and perceived meanness as mediators into separate models to investigate their mediation effects individually. Next, we examined the relative effects of these variables in a combined model. Significant interactions were probed using 1000 bootstrap samples from the original dataset and computed 95% bias-corrected bootstrap confidence intervals (95% BCCI). Model fit was assessed using conventional indices, following recommended guidelines (Hu & Bentler, 1998; Schreiber et al., 2006). These included the chi-square statistic (χ^2), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Comparative Fit Index (CFI), and Tucker–Lewis Index (TLI). Acceptable fit was defined as RMSEA and SRMR values $\leq .08$, and CFI and TLI values $\geq .90$. A non-significant χ^2 ($p > .05$) was also considered indicative of good model fit, though interpreted with caution due to its sensitivity to sample size. Additionally, we report relative fit indices—including the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and the Sample-Size Adjusted BIC (SABIC), with lower values indicating better fit—for comparisons.

## Results

Perceived intentionality

Participants accurately identified perpetrators' intentions in ~90–96% of cases involving intentional harm and in 75% of accidental harm scenarios, indicating that the experimental manipulation was successful.

Concern for victims versus perpetrators

There was a significant interaction between expressed concern for victims and perpetrators across conditions, $F(1,109) = 54.92$, $p < .001$, $\eta_p^2 = .33$. As anticipated, participants expressed significantly more concern for victims in the intentional condition, $M = 62.44$, $SE = 2.35$, 95% CI [57.78, 67.10], compared to both the accidental, $M = 44.13$, $SE = 2.51$, 95% CI [39.15, 49.11] and the control, $M = 31.82$, $SE = 2.76$, 95% CI [26.35, 37.29], conditions. In contrast, participants expressed significantly less concern for perpetrators in both the intentional, $M = 16.66$, $SE = 2.04$, 95% CI [12.61, 20.70], and control, $M = 17.21$, $SE = 2.02$, 95% CI [13.20, 21.22], conditions compared to the accidental condition, $M = 26.34$, $SE = 2.04$, 95% CI [22.30, 30.38], whereas concern for perpetrators did not differ between intentional and control conditions, $p = .793$. Additional main effect analyses showed that expressed concern for victims was significantly higher than concern for perpetrators in each condition (see Figure 2 for these contrasts). All differences were significant at $p < .001$ (uncorrected), hence remaining significant at $p < .05$ after Holm–Bonferroni correction.

Judgements of perpetrator meanness and punishment

There was a significant difference in participants' ratings of both perceived perpetrator meanness, $F(1,109) = 140.40$, $p < .001$, $\eta_p^2 = .56$, and punishment severity, $F(1,109) = 90.86$, $p < .001$, $\eta_p^2 = .45$, across conditions. Participants rated the perpetrator as significantly meaner when the aggression was intentionally directed at another person, $M = 68.96$, $SE = 2.03$, 95% CI [64.94, 72.99], or at an object,
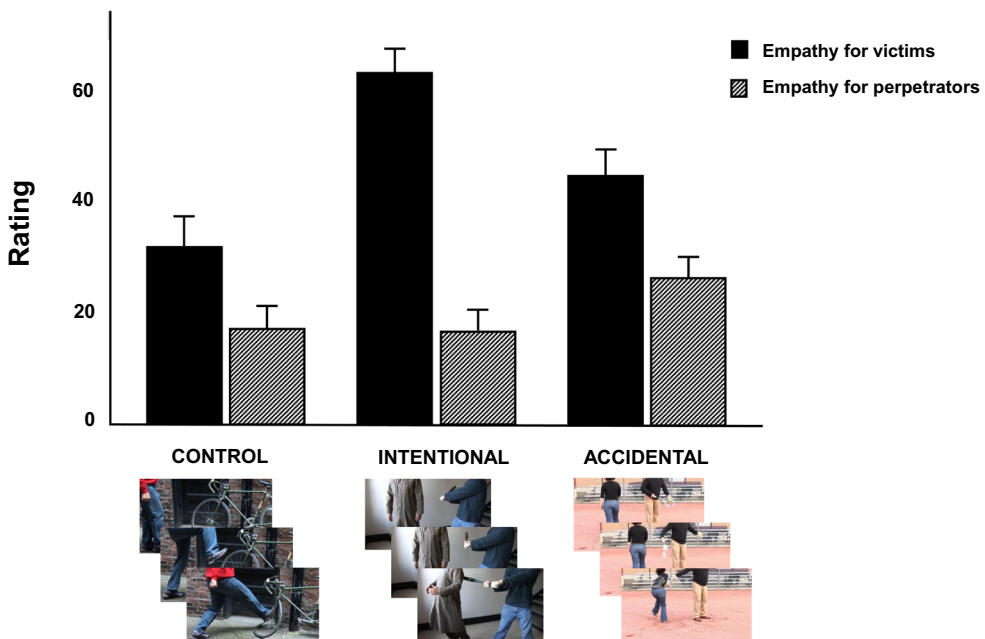
FIGURE 2 Empathic concern for victim vs perpetrator in each condition (Study 1). Error bars represent 95% confidence intervals. In all conditions, concern for victims was significantly greater than concern for perpetrators (control: $M = 31.82$, $SE = 2.76$, 95% CI [26.35, 37.29] vs. $M = 17.21$, $SE = 2.01$, 95% CI [13.20, 21.22]; intentional: $M = 62.44$, $SE = 2.35$, 95% CI [57.78, 67.10] vs. $M = 16.66$, $SE = 2.04$, 95% CI [12.61, 20.70]; accidental: $M = 44.13$, $SE = 2.51$, 95% CI [39.15, 49.11] vs. $M = 26.34$, $SE = 2.04$, 95% CI [22.30, 30.38]. All comparisons were significant at $p < .001$.

$M = 42.25$, $SE = 2.54$, 95% CI [37.23, 47.28], compared to accidental harm to another person, $M = 26.43$, $SE = 2.02$, 95% CI [22.43, 30.44]. Similarly, participants supported significantly harsher punishments for intentional harm to both a person, $M = 54.53$, $SE = 2.27$, 95% CI [50.03, 59.04], and an object, $M = 37.06$, $SE = 2.52$, 95% CI [32.08, 42.05], than for unintentional harm, $M = 21.38$, SE =1.97, 95% CI [17.46, 25.29]. Post-hoc pairwise comparisons revealed that these differences were significant at $p < .05$ after Holm–Bonferroni correction.

Direct and indirect effects of concern for victims on punishment

Table 1 provides direct and indirect effect estimates for each mediation model. Examination of individual mediation models revealed that concern for perpetrators did not significantly influence the relationship between concern for victims and punishment in neither condition. In contrast, perceived perpetrator meanness partially mediated the link between concern for victims and punishment in intentional scenarios, $\beta = 0.32$, $SE = 0.09$, 95% BCCI [0.13, 0.51], and fully mediated this relationship in accidental scenarios, $\beta = 0.40$, $SE = 0.06$, 95% BCCI [0.28, 0.54]. These effects remained when assessing concern for perpetrators and perceived meanness as mediators into combined models for each condition (see Figure 3). Fit indices indicated that both models were a good fit to our data, with relative fit indices suggesting better fit for the 'accidental' model (see Table S1).
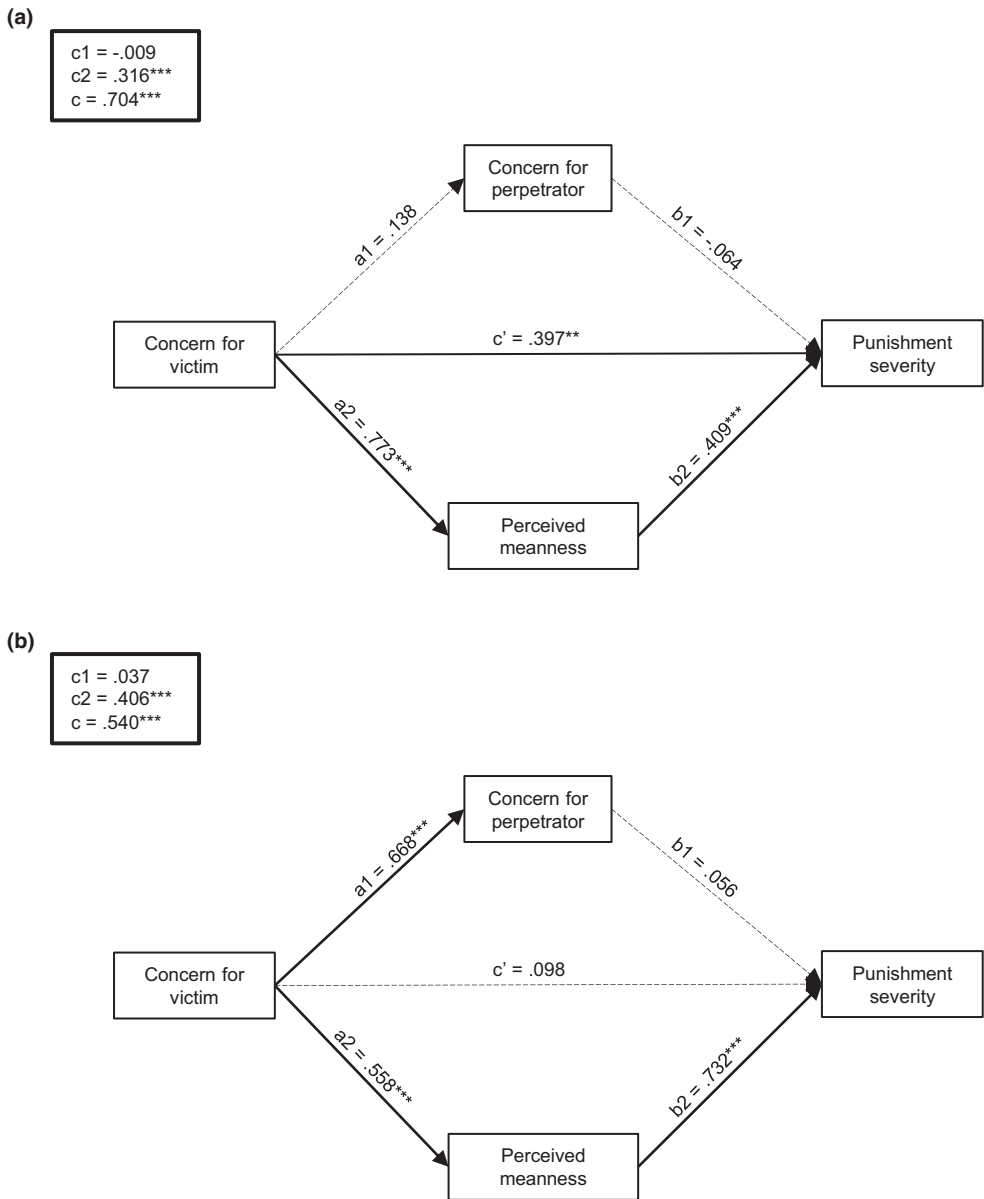
Discussion

The findings support previous research that people's judgements of perceived transgressions are significantly influenced by their perceptions of the perpetrator's intentions (Decety et al., 2012; Decety & Cowell, 2015; Mikula, 2003). Specifically, when harm is perceived as intentional, individuals tend to adopt harsher punitive attitudes towards the perpetrator (Decety et al., 2012). Furthermore, selectively

TABLE 1 Direct and indirect effect estimates of mediation models in Study 1.

	Intentional			Accidental		
	β	SE	BCCI$_{95\%}$	β	SE	BCCI$_{95\%}$
Model 1						
Total effect	.704	.052	**.593, .799**	.544	.065	**.416, .667**
Direct effect	.713	.053	**.605, .808**	.653	.128	**.410, .894**
Indirect effect	−.010	.013	−.040, .017	−.109	.094	−.294, .072
Model 2						
Total effect	.704	.055	**.591, .804**	.544	.066	**.409, .667**
Direct effect	.386	.118	**.158, .640**	.143	.082	−.020, .300
Indirect effect	.317	.094	**.135, .508**	.401	.064	**.277, .536**
Model 3						
Total effect	.704	.056	**.595, .810**	.540	.066	**.411, .672**
Direct effect	.397	.121	**.169, .637**	.097	.141	−.152, .395
Indirect effect						
Via concern for perpetrator	−.009	.014	−.035, .024	.037	.075	−.123, .176
Via perceived meanness	.316	.095	**.132, .505**	.406	.067	**.278, .535**

Note: Model 1 = Estimates through individual mediation of concern for perpetrator; Model 2 = Estimates through individual mediation of perceived perpetrator meanness; Model 3 = Estimates through the combined mediation of concern for perpetrator and perceived perpetrator meanness. Reported direct effects are estimated after accounting for indirect effects. Significant estimates (BCCI$_{95\%}$ not including 0) are in bold.

(a)

c1 = -.009
c2 = .316***
c = .704***

Concern for
perpetrator

a1 = .138

b1 = -.064

Concern for
victim

c' = .397**

Punishment
severity

a2 = .773***

b2 = .409***

Perceived
meanness

(b)

c1 = .037
c2 = .406***
c = .540***

Concern for
perpetrator

a1 = .668***

b1 = .056

Concern for
victim

c' = .098

Punishment
severity

a2 = .558***

b2 = .732***

Perceived
meanness

FIGURE 3 Path estimates in (a) intentional and (b) accidental conditions (Study 1). Mediation was assessed by examining the direct (c') and indirect effects of interpersonal callousness on aggression through concern for the perpetrator (c1) and perceived perpetrator meanness (c2). The total effect (c) was the sum of the indirect effects through the mediators and the direct effect of the predictor. *$p < .05$, **$p < .01$, ***$p < .001$.

applying empathy towards victims of aggression increased participants' tendency to impose harsher punishments on the aggressors. Notably, this effect was primarily mediated by participants' negative perceptions of the aggressor, aligning with existing literature that suggests that retaliatory responses to aggression are influenced by negative assessments of the aggressor (Carlsmith et al., 2002). In our study, this effect was particularly strong in scenarios of accidental harm, where the relationship between concern for victims and punishment severity was fully mediated by perceived perpetrator meanness.

This finding implies that people's punitive attitudes against aggressors might be more influenced by their negative perceptions of the aggressor rather than their judgement of the aggression itself or their concern for the victim. In contrast, while participants also expressed lower concern for perpetrators, this did not seem to influence their decisions to punish, contradicting our initial expectations. These effects were further examined in a follow-up study.

STUDY 2

Method

Participants

Even though the sample in Study 1 was large enough to detect the postulated effects, in this study we decided to almost triple our sample size to allow us to detect potentially small moderating effects of callous traits. A total of 310 participants ($M = 22.75$, $SD = 1.75$) were recruited via Prolific, a crowd-sourcing platform for online research (www.prolific.co). Prolific allows researchers to recruit high-quality, pre-screened participants based on demographic or psychological criteria, and it ensures ethical compensation and participant anonymity (Palan & Schitter, 2018). All participants were based in the United Kingdom and received £7 as a compensation upon successful completion of the study. After giving their informed consent, participants reported their demographic information. More than half of the sample (60.6%) identified as White, with 16.5% of the remaining participants identifying as Asian/Asian British, 13.9% as Black and 9% coming from multiple or other ethnic groups. Next, participants performed the same experimental task described in Study 1, followed by survey questionnaires. This experiment was approved by the university's ethics committee as part of a larger online study.

Assessment of callous-unemotional traits

We used the 24-item *Inventory of Callous Unemotional Traits* (ICU; Frick et al., 2003) to assess callous-unemotional traits. This questionnaire includes items denoting callousness (e.g. 'The feelings of others are unimportant to me'), disregard for the consequences of one's actions on others (e.g. 'I do not care who I hurt to get what I want'), and shallow affect (e.g. 'I do not show my emotions to others'). Responses were provided on a 4-point Likert scale from 0 (not at all true) to 3 (definitely true), and the summed score was computed for follow-up analyses—excluding items 2 ('What I think is right and wrong is different from what other people think') and 10 ('I do not let my feelings control me') due to prior research indicating higher internal consistency after removing these items, along with their low item-total correlations with the total ICU scale (Kimonis et al., 2008). After excluding these items, the remaining ICU items yielded high internal consistency, $\alpha = 0.82$.

Analysis

Initial analyses replicated the procedures described in Study 1. Additionally, we examined whether the relationship between concern for victims and punishment was moderated by levels of callous-unemotional traits in both intentional and accidental conditions. To test these effects, we conducted multiple linear regression analyses that included the main effects of concern for victims and callous-unemotional traits, as well as their interaction. Where the interaction term was significant, we used the interactions package in *R* to perform simple slopes analyses (Long & Long, 2021) estimating the effect of concern for victims on punishment at low, average, and high levels of callous-unemotional

traits. We also used the Johnson–Neyman technique to identify the range of callous-unemotional trait values for which the relationship between concern for victims and punishment was statistically significant.

Results

Perceived intentionality

Participants accurately identified intentional harm with a high success rate of 93.9%. In contrast, their accuracy in identifying unintentional harm was notably lower, at 55.5%. This discrepancy highlights the challenge participants faced in discerning the perpetrator's intentions in accidental scenarios. However, this accuracy still exceeded the pre-established 50% threshold, indicating that (for the most part) participants were still able to identify intentionality in these conditions.
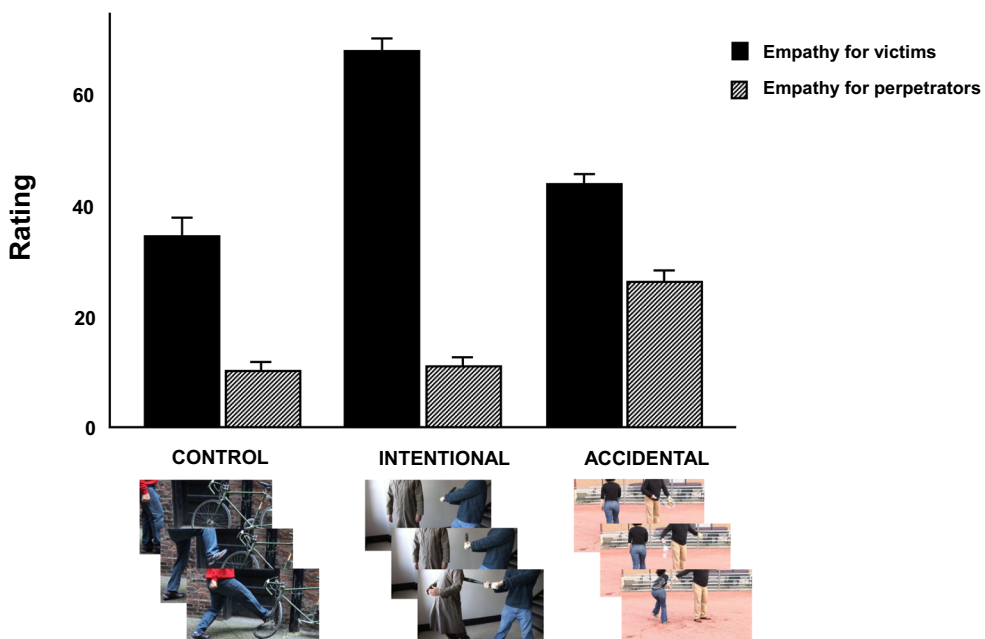
Concern for victims versus perpetrators

Consistent with our previous findings, there was a significant difference in participants' expressed concern for victims and perpetrators across conditions, $F(1, 309) = 246.46$, $p < .001$, $\eta_p^2 = .44$. However, simple main effects analyses indicated that participants expressed greater concern for victims when the transgression was intentional, $M = 66.92$, $SE = 1.28$, 95% CI [64.04, 69.44], than when it was accidental, $M = 42.76$, $SE = 1.35$, 95% CI [40.10, 45.42], or directed at an object, $M = 34.09$, $SE = 1.61$, 95% CI [30.93, 37.25]. In turn, concern for perpetrators was higher in the accidental harm condition, $M = 24.76$, $SE = 1.14$, CI 95% [23.52, 28.00.], than in both the intentional, $M = 10.66$, $SE = 0.89$, 95% CI [8.90, 12.41], and object harm, $M = 10.01$, $SE = 0.83$, 95% CI [8.38, 11.65], conditions. Further contrasts additionally showed that participants expressed significantly more concern for victims than for perpetrators in all conditions (main effects in Figure 4). All contrasts were significant at $p < .05$ after applying the Holm–Bonferroni correction, except for the difference between concern for perpetrators in the intentional versus control conditions, $p = .472$.

Moral judgements

There was a significant main effect of transgression type on both perceived meanness of the perpetrator, $F(1, 309) = 511.47$, $p < .001$, $\eta_p^2 = .62$, and punishment severity, $F(1, 309) = 371.06$, $p < .001$, $\eta_p^2 = .55$.

Participants rated the perpetrator as significantly meaner when the harm was intentional, $M = 73.61$, $SE = 1.05$, 95% CI [71.54, 75.67], compared to harm directed at an object, $M = 46.76$, $SE = 1.57$, 95% CI [43.68, 49.84] or accidental harm to a person, $M = 21.41$, $SE = 1.13$, 95% CI [19.19, 23.63]. Similarly, witnessing intentional harm to another person led to harsher punishment decisions, $M = 58.88$, $SE = 1.13$, 95% CI [56.65, 61.11], than when harm accidentally caused to another person, $M = 18.34$, $SE = 1.00$, CI95% [16.38, 20.31], or directed at an objected, $M = 37.64$, $SE = 1.41$, 95% CI [34.87, 40.42]. All comparisons were significant at $p < .001$ (uncorrected), with significance retained after Holm–Bonferroni correction.

FIGURE 4 Empathic concern for victim vs perpetrator in each condition (Study 2). Error bars represent 95% confidence intervals. Across all conditions, concern for victims was significantly higher than concern for perpetrators (Control: $M = 34.09$, $SE = 1.61$, 95% CI [30.93, 37.25] vs $M = 10.01$, $SE = 0.83$, 95% CI [8.38, 11.65]; Intentional: $M = 66.92$, $SE = 1.28$, 95% CI [64.40, 69.44] vs $M = 10.66$, $SE = 0.89$, 95% CI [8.90, 12.41]; Accidental: $M = 42.76$, $SE = 1.35$, 95% CI [40.10, 45.42] vs $M = 25.76$, $SE = 1.14$, 95% CI [23.52, 28.00]). All comparisons were significant at $p < .001$***.

Mediating effects

Table 2 provides direct and indirect effect estimates for each mediation model. When evaluating each mediator in individual models, we found that perceived meanness partially mediated the relationship between concern for victims and punishment severity in both the intentional, $\beta = 0.45$, $SE = 0.05$, 95% BCCI [0.36, 0.57], and accidental conditions, $\beta = 0.35$, $SE = 0.04$, 95% BCCI [0.27, 0.43], indicating a positive effect. In contrast, concern for perpetrators emerged as a significant mediator in the accidental condition, $\beta = -0.08$, $SE = 0.04$, 95% BCCI [−0.15, −0.01], with data suggesting a partial negative effect.

Nevertheless, when introducing both mediators into a combined model, only the effects of perceived meanness remained significant. In line with Study 1, perceived perpetrator meanness significantly mediated the effects of empathy on punishment severity in both conditions. In this case, however, we observed a partial mediation effect in the accidental condition, $\beta = 0.35$, $SE = 0.04$, 95% BCCI [0.27, 0.43], and a full mediation effect in the intentional condition, $\beta = 0.45$, $SE = 0.05$, 95% BCCI [0.36, 0.56] (see Figure 5 for an illustration). Fit indices overall indicated good fit for both models, with the 'intentional' model showing better fit. These values are reported in Table S2.

Effects of callous-unemotional traits

Lastly, we sought to explore the links between concern for victims and punishment for perpetrators through the moderation of callous-unemotional traits. As anticipated, higher concern for victims was associated with harsher punishment ratings in both the intentional, $\beta = 0.80$, $SE = 0.11$, $p < .001$, and accidental, $\beta = 0.27$, $SE = 0.10$, $p = .005$, conditions. However, in the intentional harm condition, the

TABLE 2 Direct and indirect effect estimates of mediation models in Study 2.

	Intentional			Accidental		
	β	*SE*	$BCCI_{95\%}$	β	*SE*	$BCCI_{95\%}$
Model 1						
Total effect	.614	.049	**.513, .704**	.460	.045	**.368, .549**
Direct effect	.614	.049	**.513, .704**	.542	.060	**.420, .652**
Indirect effect	−.000	.003	−.005, .005	−.082	.037	**−.154, −.007**
Model 2						
Total effect	.614	.048	**.510, .702**	.460	.046	**.372, .551**
Direct effect	.166	.075	**.022, .317**	.113	.040	**.033, .190**
Indirect effect	.447	.051	**.353, .557**	.347	.041	**.266, .426**
Model 3						
Total effect	.613	.046	**.515, .694**	.462	.047	**.370, .551**
Direct effect	.160	.070	**.019, .295**	.141	.055	**.033, .251**
Indirect effect						
via concern for perpetrator	−.000	.004	−.009, .007	−.025	.026	−.076, .026
via perceived meanness	.453	.050	**.364, .557**	.346	.041	**.269, .428**

Note: Model 1 = Estimates through individual mediation of concern for perpetrator; Model 2 = Estimates through individual mediation of perceived perpetrator meanness; Model 3 = Estimates through the combined mediation of concern for perpetrator and perceived perpetrator meanness. Reported direct effects are estimated after accounting for indirect effects. Significant estimates (95% BCCI not including 0) are in bold.

effect of victim concern on punishment severity weakened at higher levels of callous-unemotional traits, $\beta = -0.01$, $SE = 0.01$, $p = .033$. Interestingly, participants with more callous-unemotional traits also supported more severe punishment overall, $\beta = 1.09$, $SE = 0.32$, $p = .001$.
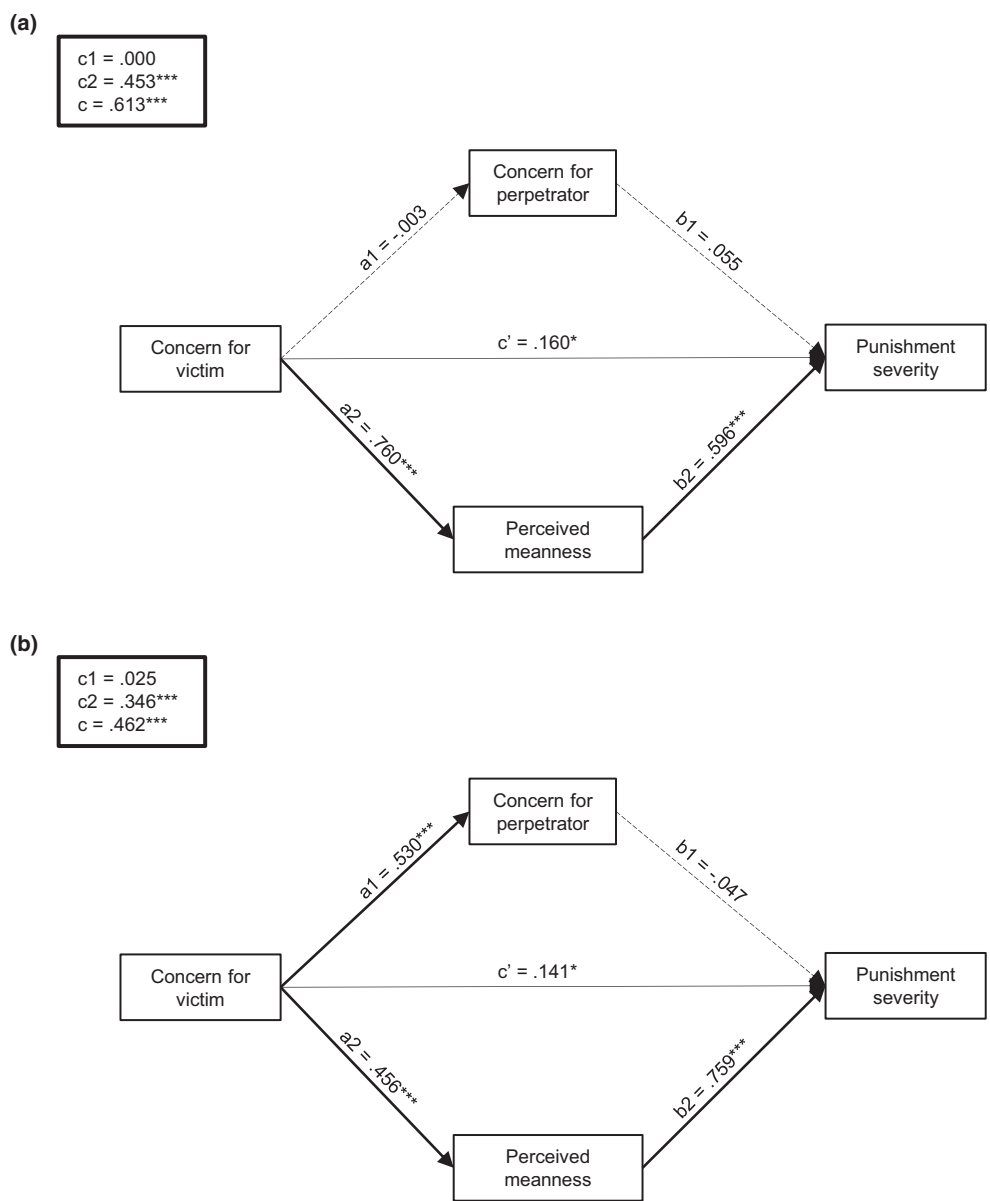
Simple slopes analyses (Figure 6) revealed that the relationship between victim concern and punishment remained significant across different levels of callous-unemotional traits (all $ps < .001$). In contrast, in the accidental harm condition, callous-unemotional traits did not significantly predict punishment severity, $\beta = 0.22$, $SE = 0.21$, $p = .314$, nor did they moderate the effect of concern for victims, $\beta = 0.00$, $SE = 0.00$, $p = .291$.

Discussion

The outcomes of this study largely replicated Study 1, demonstrating the robustness of our observations. Expanding on this, the current study further indicates that the impact of empathy on participants' punishment decisions for perceived aggression diminishes at higher levels of callous-unemotional traits. This observation supports our hypothesis that callous-unemotional traits attenuate the empathic bias in punitive decision-making.

On the other hand, however, participants with higher levels of callous-unemotional traits were also more prone to advocating for harsher penalties against aggressors. This finding challenges the notion that callous-unemotional traits might promote judicial impartiality. Instead, it suggests that individuals with these traits may be more predisposed to endorsing harsher punitive measures—which is consistent with links between callous-unemotional traits and aggression highlighted in previous research (Brugman et al., 2017; Camara et al., 2025; Frick et al., 2003; Ritchie et al., 2022). The broader implications of these findings are discussed in greater detail in the following section.

(a)

c1 = .000
c2 = .453***
c = .613***

Concern for perpetrator

a1 = -.003

b1 = .055

Concern for victim

c' = .160*

Punishment severity

a2 = .760***

b2 = .596***

Perceived meanness

(b)

c1 = .025
c2 = .346***
c = .462***

Concern for perpetrator

a1 = .530***

b1 = -.047

Concern for victim

c' = .141*

Punishment severity

a2 = .456***

b2 = .759***

Perceived meanness

F I G U R E 5   Path estimates in (a) intentional and (b) accidental conditions (Study 2). Mediation was assessed by examining the direct (c') and indirect effects of interpersonal callousness on aggression through concern for perpetrator (c1) and perceived perpetrator meanness (c2). The total effect (c) was the sum of the indirect effects through the mediators and the direct effect of the predictor. *$p < .05$, **$p < .01$, ***$p < .001$.

GENERAL DISCUSSION

Our findings align with existing research on how evaluations of transgressions are largely influenced by perceptions of the perpetrator's intentions. Consistent with the Attribution of Blame model (Shaver & Shaver, 1985), which proposes that perceived responsibility for norm violations drives moral evaluations, our results confirmed that when harm is perceived as intentional—a condition under which blame is more readily assigned—participants were more likely to endorse harsher punishments. This
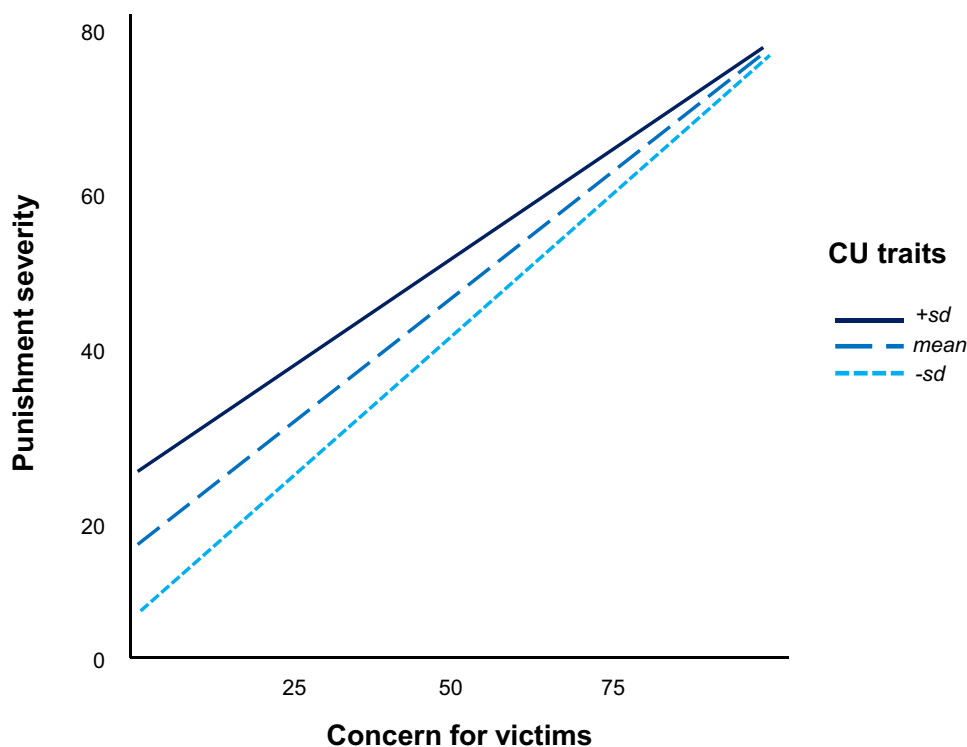
FIGURE 6 Link between punishment severity and victim bias in the intentional condition (Study 2). Effects of empathy on punishment severity at higher (+1 SD), mean, and lower (−1 SD) levels of callous-unemotional traits (CU traits).

reflects the inherent subjectivity in moral judgements, showing that the same act can be punished differently depending on the perceived intent behind it. Research additionally shows that intentional aggressions are more strongly condemned than accidental ones (Baez et al., 2014; Decety et al., 2012; Young & Saxe, 2008). Our study extends this literature by providing preliminary insights into how such evaluations relate to evaluative judgements and empathic responses towards the perpetrator.

Partially in line with our initial hypotheses, we found that participants expressed greater concern for victims than for aggressors, although their expressed concern for the perpetrator did not seem to directly influence decisions to punish. Rather, perceived meanness of the perpetrator emerged as a significant mediator in the relationship between concern for victims and punishment severity, suggesting that negative appraisals may play a more central role in justice-related decisions than empathic concern alone. This is consistent with previous research indicating that biases in punitive decision-making can be largely influenced by people's perceptions of the parties involved (e.g. Bastian et al., 2013; Viki et al., 2012). In fact, it is plausible that perceived perpetrator meanness influences empathic responses to the aggressor as well. That is, seeing the perpetrator as 'mean' may serve to justify the withholding of empathy and reinforce punishment (e.g. Bastian et al., 2013; Harris & Rice, 2006; Osofsky et al., 2005). In this view, one potential avenue for follow-up research could be to further examine whether the link between concern for perpetrators and punishment severity might be mediated by perceived perpetrator meanness.

Furthermore, the current work contributes to the literature by examining how individual differences in callous-unemotional traits influence these empathic processes and their relationship to punitive attitudes, opening new avenues for further exploration. Our findings indicate that individuals with more callous-unemotional traits exhibit lower concern for victims and a weaker link between victim-focused empathy and punishment yet still tend to favour harsher punishments overall. Interestingly, this pattern aligns with findings from studies of patients with the behavioural variant of frontotemporal dementia

(bvFTD)—a neurodegenerative condition marked by poor affect, diminished empathy and impaired social cognition (Piguet et al., 2011; Rascovsky et al., 2011). Patients with bvFTD have been shown to rate punishments as more severe, regardless of whether harm was inflicted intentionally or accidentally (Baez et al., 2014), a pattern that mirrors the reduced concern and elevated punishment ratings observed at higher levels of callous-unemotional traits in our study. These convergences support the view that traits associated with emotional detachment and reduced affective resonance may reduce the impact of perceived intentionality on moral evaluations.

Supporting this interpretation, our data further showed that in scenarios involving accidental harm, callous-unemotional traits did not moderate the relationship between concern for victims and punishment, even though they were correlated with increased punishment severity in these cases. This could suggest that punitive decisions among individuals with higher levels of callous-unemotional traits may be less guided by empathic sensitivity or moral evaluation, and more by a rigid or indiscriminate orientation towards punishment—consistent with prior research indicating that callous-unemotional traits confer unique risks for more instrumental or severe forms of aggression (e.g. Brugman et al., 2017; Camara et al., 2025; Frick et al., 2003). These possibilities gain special relevance in legal and forensic contexts, where individual traits may predispose certain decision-makers (e.g. jurors) to punitive bias, regardless of contextual factors like intent or remorse. They also underscore the importance of considering personality-driven variations in justice motivation, which may help inform targeted interventions or training programs aimed at promoting fairness and reducing bias in evaluative judgements.

Limitations

Altogether, our findings highlight the nuanced role of perceptions of both aggression and the aggressor in shaping punishment decisions, suggesting that punitive attitudes are influenced by a combination of empathy, cognitive evaluation and context. However, these interpretations should be approached with caution due to the study limitations. One key issue lies in the underlying assumption that emotion and partiality are intrinsically linked, suggesting that impartiality is more unemotional or rational. While the ideal of judicial dispassion is deeply ingrained in legal theory and popular opinion, recent arguments (e.g. Maroney, 2024) suggest that emotions can be effectively managed and do not inevitably lead to bias or partiality. Relatedly, our analysis of fairness and justice has primarily focused on emotional influences, yet these concepts encompass a broader range of factors. For instance, elements such as social identity, familiarity and other contextual variables might also predispose individuals to certain biases in their moral judgements (e.g. Abbink & Harris, 2019; Schiller et al., 2014). Consequently, the scope of our analysis is limited, and future research should consider these interactive factors to provide a more comprehensive understanding of the dynamics underlying punitive decision-making.

In addition, the lack of a consistent definition of empathy poses challenges in generalizing our findings to its various dimensions. As noted by Brown (2012, p. 386): 'Because the definitions and understandings of empathy we work with are generally positioned in the realm of the abstract, then empathy 'in the field' and 'on the ground' may require alternative dimensions in its articulation'. In other words, given the multifaceted nature of empathy, reliance on one conceptualization may risk oversimplifying the role empathy plays on behavioural outcomes, as it neglects the contextual and situational factors that influence how empathy is expressed in everyday social behaviour. As such, it is essential to clarify which aspect of empathy is under investigation—be it cognitive or affective—and to consider how these facets manifest differently across various social contexts (Vachon & Lynam, 2016). This distinction is crucial because each facet of empathy might interact with moral judgement and punitive decision-making in unique ways. Although fully resolving these definitional ambiguities is beyond the scope of our current work, it highlights an important consideration that should be explored in future studies looking into the role of empathy in punishment.

Moreover, the use of animated images to elicit empathy is another key limitation in the current research. While animated stimuli may offer a more dynamic representation compared to static images

(Decety et al., 2012), they cannot fully capture the complexity and nuance of real-life situations. In everyday scenarios, moral judgements are often influenced by the contextual details surrounding the actions of the perpetrator and victim (Jin & Peng, 2021). By omitting such contextual information, our study may have limited the scope of participants' responses, preventing a more comprehensive understanding of the factors that influence punitive decision-making. Future research that incorporates more detailed background context may help capture a broader range of reactions and provide a more accurate reflection of real-world moral judgements.

Finally, while our sample was designed to examine patterns in a normative population, our exploration of callous-unemotional traits would benefit from studying a sample with a broader or more pronounced distribution of these traits. Individuals with higher levels of callous-unemotional traits—such as those classified as 'psychopaths'—may exhibit more distinct patterns of empathy and punishment, offering a deeper understanding of their role in justice-related decisions. This approach could provide a more nuanced account of how specific socio-affective traits shape punitive attitudes and behaviours.

CONCLUSION

Overall, the common thread across our findings is that people's punishment decision-making is influenced by their perceptions of both the aggression and the aggressor, although the latter seems to have a greater effect. While punishment severity significantly increased for intentional transgressions, negative perceptions of the perpetrator influenced punishment regardless of the intentionality of the transgression itself. This suggests that punitive responses are not solely guided by principles of justice—such as proportionality or the need for correction—but are also susceptible to biases rooted in how we perceive those we judge. Notably, the tendency for individuals with more callous-unemotional traits to endorse harsher punishment regardless of perceived intentionality suggests that punishment may also reflect something deeper in an individual's moral compass, such as a greater predisposition to engage in interpersonal harm (especially if there is no motivation for retribution or vengeance). These reflections provide a basis for future research to explore the broader implications of callous-unemotional traits in moral judgement and punitive decision-making.

AUTHOR CONTRIBUTIONS
Célia F. Camara: Conceptualization; investigation; writing – original draft; methodology; formal analysis; project administration; visualization; data curation. **Alejandra Sel:** Funding acquisition; supervision; writing – review and editing. **Paul H. P. Hanel:** Conceptualization; validation; writing – review and editing; supervision; project administration.

CONFLICT OF INTEREST STATEMENT
The authors report no conflicts of interest related to this work.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in Open science framework at https://osf.io/bfmnt/?view_only=180a8ea9aea140b69f97af6b4b238ffc.

ORCID
Célia F. Camara https://orcid.org/0009-0004-5204-9161

REFERENCES

Abbink, K., & Harris, D. (2019). In-group favouritism and out-group discrimination in naturally occurring groups. *PLoS One*, *14*(9), e0221616. https://doi.org/10.1371/journal.pone.0221616

Ames, D. L., Jenkins, A. C., Banaji, M. R., & Mitchell, J. P. (2008). Taking another person's perspective increases self-referential neural processing. *Psychological Science*, *19*(7), 642–644. https://doi.org/10.1111/j.1467-9280.2008.02135.x

Baez, S., Herrera, E., García, A. M., Manes, F., Young, L., & Ibáñez, A. (2017). Outcome-oriented moral evaluation in terrorists. *Nature Human Behaviour*, *1*(6), 118.

Baez, S., Manes, F., Huepe, D., Torralva, T., Fiorentino, N., Richter, F., Huepe-Artigas, D., Ferrari, J., Montañes, P., & Reyes, P. (2014). Primary empathy deficits in frontotemporal dementia. *Frontiers in Aging Neuroscience*, *6*, 262.

Barchia, K., & Bussey, K. (2011). Individual and collective social cognitive influences on peer aggression: Exploring the contribution of aggression efficacy, moral disengagement, and collective efficacy. *Aggressive Behavior*, *37*(2), 107–120. https://doi.org/10.1002/ab.20375

Bastian, B., Denson, T. F., & Haslam, N. (2013). The roles of dehumanization and moral outrage in retributive justice. *PLoS One*, *8*(4), e61842. https://doi.org/10.1371/journal.pone.0061842

Bastian, B., Laham, S. M., Wilson, S., Haslam, N., & Koval, P. (2011). Blaming, praising, and protecting our humanity: The implications of everyday dehumanization for judgments of moral status. *British Journal of Social Psychology*, *50*(3), 469–483. https://doi.org/10.1348/014466610X521383

Batson, C. D. (1997). Self–other merging and the empathy–altruism hypothesis: Reply to Neuberg et al. *(1997)*.

Batson, C. D., O'Quin, K., Fultz, J., Vanderplas, M., & Isen, A. M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *Journal of Personality and Social Psychology*, *45*(3), 706–718. https://doi.org/10.1037/0022-3514.45.3.706

Brown, M. (2012). Empathy and punishment. *Punishment & Society*, *14*(4), 383–401. https://doi.org/10.1177/1462474512452507

Brugman, S., Cornet, L. J. M., Smeijers, D., Smeets, K., Oostermeijer, S., Buitelaar, J. K., Verkes, R., Lobbestael, J., De Kogel, C. H., & Jansen, L. M. C. (2017). Examining the reactive proactive questionnaire in adults in forensic and non-forensic settings: A variable- and person-based approach. *Aggressive Behavior*, *43*(2), 155–162. https://doi.org/10.1002/ab.21671

Camara, C. F., Sel, A., de Klerk, C. C., & Hanel, P. H. (2025). Direct and indirect effects of interpersonal callousness on aggression through empathy and moral disengagement. *Personality and Individual Differences*, *232*, 112836. https://doi.org/10.1016/j.paid.2024.112836

Campos, C., Pasion, R., Azeredo, A., Ramião, E., Mazer, P., Macedo, I., & Barbosa, F. (2022). Refining the link between psychopathy, antisocial behavior, and empathy: A meta-analytical approach across different conceptual frameworks. *Clinical Psychology Review*, *94*, 102145. https://doi.org/10.1016/j.cpr.2022.102145

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

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

Čehajić, S., Brown, R., & González, R. (2009). What do I care? Perceived ingroup responsibility and dehumanization as predictors of empathy felt for the victim group. *Group Processes & Intergroup Relations*, *12*(6), 715–729. https://doi.org/10.1177/1368430209347727

Cikara, M., & Fiske, S. T. (2013). Their pain, our pleasure: Stereotype content and schadenfreude. *Annals of the New York Academy of Sciences*, *1299*(1), 52–59.

Cima, M., Tonnaer, F., & Hauser, M. D. (2010). Psychopaths know right from wrong but don't care. *Social Cognitive and Affective Neuroscience*, *5*(1), 59–67. https://doi.org/10.1093/scan/nsp051

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

Decety, J. (2010). The neurodevelopment of empathy in humans. *Developmental Neuroscience*, *32*(4), 257–267. https://doi.org/10.1159/000317771

Decety, J., & Cowell, J. M. (2015). Empathy, justice, and moral behavior. *AJOB Neuroscience*, *6*(3), 3–14. https://doi.org/10.1080/21507740.2015.1047055

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100. https://doi.org/10.1177/1534582304267187

Decety, J., Michalska, K. J., & Kinzler, K. D. (2012). The contribution of emotion and cognition to moral sensitivity: A neurodevelopmental study. *Cerebral Cortex*, *22*(1), 209–220. https://doi.org/10.1093/cercor/bhr111

Decety, J., & Yoder, K. J. (2016). Empathy and motivation for justice: Cognitive empathy and concern, but not emotional empathy, predict sensitivity to injustice for others. *Social Neuroscience*, *11*(1), 1–14. https://doi.org/10.1080/17470919.2015.1029593

Dutton, K. (2012). *The wisdom of psychopaths*. Random House.

Feather, N. T. (1994). Human values and their relation to justice. *Journal of Social Issues*, *50*(4), 129–151. https://doi.org/10.1111/j.1540-4560.1994.tb01201.x

Feather, N. T., & Dawson, S. (1998). Judging deservingness and affect in relation to another's employment or unemployment: A test of a justice model. *European Journal of Social Psychology*, *28*(3), 361–381. https://doi.org/10.1002/(SICI)1099-0992(199805/06)28:3<361::AID-EJSP867>3.0.CO;2-E

Fragkaki, I., Cima, M., & Meesters, C. (2016). The association between callous–unemotional traits, externalizing problems, and gender in predicting cognitive and affective morality judgments in adolescence. *Journal of Youth and Adolescence*, *45*(9), 1917–1930. https://doi.org/10.1007/s10964-016-0527-x

Frick, P. J., Cornell, A. H., Barry, C. T., Bodin, S. D., & Dane, H. E. (2003). Callous-unemotional traits and conduct problems in the prediction of conduct problem severity, aggression, and self-report of delinquency. *Journal of Abnormal Child Psychology*, *31*, 457–470.

Gini, G., Pozzoli, T., & Bussey, K. (2015). TThe role of individual and collective moral disengagement in peer aggression and bystanding: A multilevel analysis. *Journal of Abnormal Child Psychology*, *43*(3), 441–452. https://doi.org/10.1007/s10802-014-9920-7

Hall, J. R., & Benning, S. D. (2006). The 'successful' psychopath: Adaptive and subclinical manifestations of psychopathy in the general population. In C. J. Patrick (Ed.), *Handbook of psychopathy* (pp. 459–478). The Guilford Press.

Harris, G. T., & Rice, M. E. (2006). Treatment of psychopathy: A review of empirical findings. In C. J. Patrick (Ed.), *Handbook of psychopathy* (pp. 555–572). The Guilford Press.

Haslam, S. A., O'Brien, A., Jetten, J., Vormedal, K., & Penna, S. (2005). Taking the strain: Social identity, social support, and the experience of stress. *British Journal of Social Psychology*, *44*(3), 355–370. https://doi.org/10.1348/014466605X37468

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453.

Jin, W. Y., & Peng, M. (2021). The effects of social perception on moral judgment. *Frontiers in Psychology*, *11*, 557216. https://doi.org/10.3389/fpsyg.2020.557216

Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Munoz, L. C., Aucoin, K. J., & Morris, A. S. (2008). Assessing callous–unemotional traits in adolescent offenders: Validation of the inventory of callous–unemotional traits. *Psychopathic Traits and Risk Assessment in Children and Adolescents*, *31*(3), 241–252. https://doi.org/10.1016/j.ijlp.2008.04.002

Koenigs, M., Kruepke, M., Zeier, J., & Newman, J. P. (2012). Utilitarian moral judgment in psychopathy. *Social Cognitive and Affective Neuroscience*, *7*(6), 708–714. https://doi.org/10.1093/scan/nsr048

Kohlberg, L. (1981). *The philosophy of moral development: Moral stages and the idea of justice*. Harper & Row.

Lalonde, R. N., & Silverman, R. A. (1994). Behavioral preferences in response to social injustice: The effects of group permeability and social identity salience. *Journal of Personality and Social Psychology*, *66*(1), 78–85. https://doi.org/10.1037/0022-3514.66.1.78

Levy, D. A. (2022). Optimizing the social utility of judicial punishment: An evolutionary biology and neuroscience perspective. *Frontiers in Human Neuroscience*, *16*, 967090. https://doi.org/10.3389/fnhum.2022.967090

Lin, Z., Cui, F., Wu, Y., & Wei, Q. (2024). The effect of wrongdoer's status on observer punishment recommendations: The mediating role of envy and the moderating role of belief in a just world. *Frontiers in Psychology*, *15*, 1227961. https://doi.org/10.3389/fpsyg.2024.1227961

Lockwood, P. L., Sebastian, C. L., McCrory, E. J., Hyde, Z. H., Gu, X., De Brito, S. A., & Viding, E. (2013). Association of callous traits with reduced neural response to others' pain in children with conduct problems. *Current Biology*, *23*(10), 901–905. https://doi.org/10.1016/j.cub.2013.04.018

Long, J. A., & Long, M. (2021). *Package 'interactions': Comprehensive, user-friendly toolkit for probing interactions*. [R package manual] https://cran.r-project.org/web/packages/interactions/interactions.pdf

Lu, T., & McKeown, S. (2018). The effects of empathy, perceived injustice and group identity on altruistic preferences: Towards compensation or punishment. *Journal of Applied Social Psychology*, *48*(12), 683–691. https://doi.org/10.1111/jasp.12532

Maroney, T. A. (2024). Law and emotion. In G. B. Spencer & J. Smith (Eds.), *Research handbook on law and psychology* (pp. 64–82). Edward Elgar Publishing.

Mikula, G. (2003). Testing an attribution-of-blame model of judgments of injustice. *European Journal of Social Psychology*, *33*(6), 793–811. https://doi.org/10.1002/ejsp.184

Neuberg, S. L., Cialdini, R. B., Brown, S. L., Luce, C., Sagarin, B. J., & Lewis, B. P. (1997). Does empathy lead to anything more than superficial helping? Comment on Batson et al. (1997). *Personality and Social Psychology Bulletin*, *23*(2), 119–124.

Osofsky, M. J., Bandura, A., & Zimbardo, P. G. (2005). The role of moral disengagement in the execution process. *Law and Human Behavior*, *29*(4), 371–393. https://doi.org/10.1007/s10979-005-4930-1

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27.

Perren, S., & Gutzwiller-Helfenfinger, E. (2012). Cyberbullying and traditional bullying in adolescence: Differential roles of moral disengagement, moral emotions, and moral values. *European Journal of Developmental Psychology*, *9*(2), 195–209. https://doi.org/10.1080/17405629.2011.643168

Piaget, J. (1932). *The moral judgment of the. Chi1d*. McGraw-Hil1 Book Co.

Piguet, O., Hornberger, M., Mioshi, E., & Hodges, J. R. (2011). Behavioural-variant frontotemporal dementia: Diagnosis, clinical staging, and management. *The Lancet Neurology*, *10*(2), 162–172.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Rascovsky, K., Hodges, J. R., Knopman, D., Mendez, M. F., Kramer, J. H., Neuhaus, J., Van Swieten, J. C., Seelaar, H., Dopper, E. G., & Onyike, C. U. (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*, *134*(9), 2456–2477.

Ritchie, M. B., Neufeld, R. W. J., Yoon, M., Li, A., & Mitchell, D. G. V. (2022). Predicting youth aggression with empathy and callous unemotional traits: A meta-analytic review. *Clinical Psychology Review*, *98*, 102186. https://doi.org/10.1016/j.cpr.2022.102186

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), i02. https://doi.org/10.18637/jss.v048.i02

Ruby, P., & Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *Journal of Cognitive Neuroscience*, *16*(6), 988–999. https://doi.org/10.1162/0898929041502661

Santamaría-García, H., Baez, S., Reyes, P., Santamaría-García, J. A., Santacruz-Escudero, J. M., Matallana, D., Arévalo, A., Sigman, M., García, A. M., & Ibáñez, A. (2017). A lesion model of envy and schadenfreude: Legal, deservingness and moral dimensions as revealed by neurodegeneration. *Brain*, *140*(12), 3357–3377.

Schiller, B., Baumgartner, T., & Knoch, D. (2014). Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination. *Evolution and Human Behavior*, *35*(3), 169–175. https://doi.org/10.1016/j.evolhumbehav.2013.12.006

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323–338.

Shaver, K. G., & Shaver, K. G. (1985). A theory of blame. The attribution of blame: Causality, responsibility, and blameworthiness.

Shichman, R., & Weiss, A. (2022). Sending a message or seeking retributive justice? An experimental investigation of why people punish. *Journal of Behavioral and Experimental Economics*, *100*, 101918. https://doi.org/10.1016/j.socec.2022.101918

Simas, E. N., Clifford, S., & Kirkland, J. H. (2020). How empathic concern fuels political polarization. *American Political Science Review*, *114*(1), 258–269.

Tarrant, M., Branscombe, N. R., Warner, R. H., & Weston, D. (2012). Social identity and perceptions of torture: It's moral when we do it. *Journal of Experimental Social Psychology*, *48*(2), 513–518.

Tarrant, M., Dazeley, S., & Cottom, T. (2009). Social categorization and empathy for outgroup members. *British Journal of Social Psychology*, *48*(3), 427–446.

Tyler, T. R., & Blader, S. L. (2003). The group engagement model: Procedural justice, social identity, and cooperative behavior. *Personality and Social Psychology Review*, *7*(4), 349–361.

Underwood, B., & Moore, B. (1982). Perspective-taking and altruism. *Psychological Bulletin*, *91*(1), 143–173. https://doi.org/10.1037/0033-2909.91.1.143

Vachon, D. D., & Lynam, D. R. (2016). Fixing the problem with empathy: Development and validation of the affective and cognitive measure of empathy. *Assessment*, *23*(2), 135–149. https://doi.org/10.1177/1073191114567941

Vanman, E. J. (2016). The role of empathy in intergroup relations. *Current Opinion in Psychology*, *11*, 59–63.

Vasconcelos, M., Viding, E., Sebastian, C. L., Faria, S., Almeida, P. R., Gonçalves, Ó. F., Gonçalves, R. A., Sampaio, A., & Seara-Cardoso, A. (2021). Callous-unemotional traits moderate anticipated guilt and wrongness judgments to everyday moral transgressions in adolescents. *Frontiers in Psychiatry*, *12*, 328.

Viki, G. T., Fullerton, I., Raggett, H., Tait, F., & Wiltshire, S. (2012). The role of dehumanization in attitudes toward the social exclusion and rehabilitation of sex offenders. *Journal of Applied Social Psychology*, *42*(10), 2349–2367. https://doi.org/10.1111/j.1559-1816.2012.00944.x

Vitaglione, G. D., & Barnett, M. A. (2003). Assessing a new dimension of empathy: Empathic anger as a predictor of helping and punishing desires. *Motivation and Emotion*, *27*, 301–325.

Weiner, B. (1995). Inferences of responsibility and social motivation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. *27*, pp. 1–47). Academic Press. https://doi.org/10.1016/S0065-2601(08)60402-5

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage*, *40*(4), 1912–1920.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Camara, C. F., Sel, A., & Hanel, P. H. P. (2025). When empathy leads to aggression: The effects of empathy on punitive attitudes towards aggressors. *British Journal of Social Psychology*, *64*, e12907. https://doi.org/10.1111/bjso.12907