The Influence of Offender Identifiability in Second- and Third-Party Punishment

[Registered Report, Stage 1, Version 2]

^Theresa Blanke

ORCID: 0000-0002-4434-3289

Department Psychology, Ludwig-Maximilians-Universität München

t.blanke@campus.lmu.de / theresa.blanke@gmail.com

Mathias Twardawski

ORCID: 0000-0003-0543-277X

Department Psychology, Ludwig-Maximilians-Universität München

m.twardawski@psy.lmu.de

^ Corresponding author

Author bios:

Theresa Blanke is a master's student at the Ludwig-Maximilians-Universität München. She will conduct this study as her master's thesis.

Mathias Twardawski is a research associate at the chair of Social Psychology, Ludwig-Maximilians-Universität München. His research focuses on the treatment of interpersonal misbehavior.

Declaration of Conflict of Interest:

The author(s) declare no potential conflicts of interest concerning the authorship and/or publication of this article.

Financial disclosure/funding:

The authors received no financial support for the research and/or authorship of this article. [open to revision following in-principle acceptance.]

Authorship declaration:

Theresa Blanke will conduct this study as part of her master thesis in Psychology. Mathias Twardawski guided and supervised each step of the study report. [Open to revision following in-principle acceptance.]

Corresponding author

Theresa Blanke, Department of Psychology, Ludwig-Maximilian Universität München, Munich GER; theresa.blanke@gmail.com; 0000-0002-4434-3289

Open practice statement

The authors agree to share the anonymized raw data and digital study materials (including stimuli, experiment code, and analysis code) for all published results.

Abstract

The identifiability effect describes humans' tendency to exhibit different emotional and behavioral responses toward identified as compared to anonymous individuals. This phenomenon has been extensively studied within the identifiable victim effect, ~~showing that~~hypothesizing that people display more helping behavior when faced with an identified as compared to an anonymous individual in need. By contrast, research on the influence of offender identifiability on the perception and treatment of offenders is relatively scarce and has found mixed results. Specifically, some studies show negative consequences of offender identifiability for offenders (e.g., stronger punishment); whereas~~,~~ other studies found positive consequences for offenders (e.g., weaker punishment). In the present study, we aim to investigate whether the consequences of identifiability (vs. anonymity) of the offender may vary depending on the punisher's role in the initial offense. We hypothesize that offender identifiability leads to stronger punishment for offenders when punishment is imposed by the initial victim. By contrast, offender identifiability should lead ~~but~~ to weaker punishment when punishment is imposed by an uninvolved third party. To investigate this, we propose an online vignette study in which participants read about an offense (a) in which they were either the affected victim themselves (second party) or uninvolved and only learned about the offense (third party) and (b) in which the offender is either identified or unidentified. Following this, participants will indicate their intentions to punish, as well as their levels of empathy, moral outrage, and attribution of blame towards the offender.

Keywords: offender identifiability, second-party punishment, third-party punishment, offender empathy, moral outrage, offender blaming

The Influence of Offender Identifiability in Second- and Third-Party Punishment

Cases of victimization often garner significant attention in the public sphere, frequently becoming focal points in media discourse. However, these cases are particularly delicate, involving highly vulnerable individuals in highly sensitive situations. Therefore, protecting both victims and offenders is of paramount importance. In Germany, for example, the anonymity of both victims and offenders must be preserved in media crime reporting and court proceedings to protect their personal rights (Hooffacker & Meier, 2017). Victims receive special protection and their identity is generally not disclosed unless consent is given or they are considered public figures. The identity of offenders (e.g., name and photo), by contrast, can or should be disclosed under certain circumstances, especially when there is a legitimate public interest that outweighs the individual's protective interests. Disclosing the identity of victims or offenders may, however, change people's attitudes and subsequent behavior toward them (Lewinsohn-Zamir et al., 2017; Pałka et al., 2023), ultimately influencing court decisions and judicial sentencing (Hough et al., 2003; Myers & Greene, 2004; Roberts & Hough, 2005). Given the potentially far-reaching impact of such seemingly trivial details, like disclosing the identities of victims and offenders, it is essential to understand its actual consequences and accompanying psychological mechanisms.

Previous research has extensively studied the impact of the so-called “identifiability effect” on the perception and public support for victims (e.g., Genevsky et al., 2013; Lee & Feeley, 2016). By contrast, the influence of identifiability (vs. anonymity) of offenders on other individuals’ emotions, attitudes, and behavior has received limited attention. Moreover, existing research has provided mixed results as to whether offender identifiability has positive (Barak-Corren & Lewinsohn-Zamir, 2019; Baron & Ritov, 2009; Kogut, 2011b; Lewinsohn-Zamir et al., 2017) or negative consequences (Kogut, 2011b; Small & Loewenstein, 2005) for offenders. The current study aims to contribute to this literature and particularly suggests that the influence of offender identifiability on the perception and treatment of the offender

depends on the punisher's role in the initial offense. More specifically, we test the hypothesis that offender identifiability has negative consequences for the offender (e.g., stronger punishment intentions) when the situation is evaluated by the initial victim; conversely, offender identifiability should have positive consequences for the offender (e.g., weaker punishment intentions) when the situation is evaluated by a third party that was unaffected by the initial offense.

The Identifiability Effect

Research on the identifiability effect originated in 1968 when Schelling first described the phenomenon that the death of a single individual life triggers stronger emotional reactions and helping behavior than the death of an anonymous, so-called "statistical" life. In essence, Schelling summarized his idea as "the more we know, the more we care" (Schelling, 1968). In line with this reasoning, ~~decades of research~~ a large body of research has tested the hypothesis ~~suggest~~ that the mere identification of a victim by "irrelevant information" such as name, age, or a picture leads to stronger emotional reactions (e.g., empathy) and willingness to help (e.g., donation behavior) compared to cases where the victim remains anonymous (e.g., Hou et al., 2023; Kogut & Ritov, 2005a; Lee & Feeley, 2018; Small et al., 2007)[1]. One theoretical explanation that has been proposed for this effect is that, by making the individual more concrete and increasing their psychological closeness (Mentovich et al., 2016; Pałka et al., 2023), identifiability amplifies positive and benevolent emotions, and, ultimately, increases positive behavioral intentions concerning the victim, such as helping behavior (Kogut et al., 2018; Kogut & Ritov, 2005a; Oppenheimer & Olivola, 2011). Of note, recent studies have raised questions about the robustness and strength of the victim identifiability effect. Maier

[1] The identifiability effect refers to the impact of providing identifying information – often unrelated to the core issue – about a single individual. This effect needs to be distinguished from related concepts such as *singularity* or *scope insensitivity*, which examine whether a single identified individual elicits stronger (emotional) reactions compared to a group or a larger number of (identified or unidentified) individuals (Kogut & Ritov, 2005b, 2005a; Maier et al., 2024; Moche et al., 2022; Sudhir et al., 2016). In the present research, we specifically investigate the impact of having identifying information versus no identifying information about one individual.

and colleagues (2024), for example, conducted a re-analysis of meta-analytic evidence on the identifiable victim effect and reported moderate evidence of publication bias, and strong evidence for the absence of the average identified victim effect. Additionally, several recent studies have failed to replicate the victim identifiability effect (e.g., Majumder et al., 2022; Moche & Västfjäll, 2021; Vu et al., 2024; Wiss et al., 2015), further complicating the understanding of its strength and the conditions under which it operated in the literature. Given this mixed evidence, the extent and boundary conditions of the victim identifiability effect remain uncertain. ~~Although this effect appears to be highly context-dependent and has recently failed to replicate, a great body of evidence indicates that, in general, it is advantageous for victims to be identified .~~

~~By contrast~~In contrast to the extensive research on the victim identifiability effect, ~~evidence on~~ the effects of *offender* identifiability on individuals' perceptions of and responses to an offense ~~is strikingly scarce~~ received relatively little attention and, thus, remain insufficiently explored. ~~and yielded~~ Moreover, the few studies testing this effect yielded remarkably inconsistent results (as we will outline below). One of the first attempts to extend the victim identifiability effect and to test how the identification of offenders influences people's emotional responses and behavior toward the offender has been provided by Small and Loewenstein (2005). The authors suggested that identification generally intensifies prevailing emotional and moral reactions toward an individual rather than only amplifying positive and benevolent responses (see also Loewenstein et al., 2005). To test this reasoning, Small and Loewenstein (2005) employed an economic game in which participants became victims of injustice due to the unfair behavior of either an identified or unidentified teammate (the offender). They could then decide how much of their own money they wanted to invest in order to reduce the offender's payout and, thus, punish them. ~~in which participants could choose to either share their money with the group (contributors), which benefited everyone, or keep their money for themselves (non-contributors), which increased their own profit but~~

~~reduced the profits of other contributing participants. In a second round of the game, contributors could then decide whether and to what extent they would like to invest money to punish non-contributors. Importantly, before punishment, non-contributors were either identified or not.~~ In line with their hypothesis, results showed that feelings of anger and blame were amplified and, ultimately, led to stronger punitive actions against the offender ~~(i.e., the non-contributor)~~ when the offender was identified, compared to when the offender remained unidentified. This research led to the general belief that offender identifiability enhances individuals' tendency to punish the offender (Baron & Ritov, 2009; Bystranowski et al., 2022; Feng et al., 2023; Hsu, 2008; Sah & Loewenstein, 2012; Żuradzki, 2019).

However, other research paints a completely different picture and suggests that the effects are not quite that simple. Baron and Ritov (2009, Study 8), for example, also expected that offender identifiability increases punishment intentions. However, their data revealed patterns that were in the direct opposite direction: People expressed *weaker* intentions to punish identified compared to unidentified offenders. Similarly, Lewinsohn-Zamir and colleagues (2017) found that in cases of minor offenses (e.g., littering in the park or breach of contract), offender identifiability made people punish more *leniently*. Barak-Corren and Lewinsohn-Zamir (2019) investigated the effect of victim and offender identifiability on third-party perceptions of credibility, morality, blameworthiness, and responsibility of both victims and offenders in the case of (accused) sexual harassment. Again, data revealed that offender identifiability was advantageous for offenders: Participants evaluated identified (vs. unidentified) offenders more positively and were less supportive of measures against the offender (e.g., disciplinary actions, filing criminal charges). This research would, thus, suggest that offender identifiability reduces (rather than enhances) individuals' tendency to punish the offender.

In sum, previous studies show mixed results regarding whether identifying an offender leads to more benevolent versus malevolent emotions and behavior toward them. One aspect that has a fundamental impact on people's punishment in general and that may also help explain~~ing~~ these mixed results is the role of the punisher in the initial offense.

Second- and Third-Party Punishment

Punishment of the offender can most prominently be imposed by victims of the offense. This has been termed second-party punishment (e.g., Mischkowski et al., 2018). Sometimes, however, punishment is enforced by individuals who were unaffected by the initial offense and, thus, were only in an observer role. This has been termed third-party punishment (e.g., Henrich et al., 2006). There is a great body of literature investigating the parallels and differences between these two positions on punishment decisions (Pedersen et al., 2018; Przepiorka & Liebe, 2016; Twardawski et al., 2022). This research, for example, suggests that second parties tend to punish more strongly than third parties (Civai et al., 2019; Fehr & Fischbacher, 2004; Hechler & Kessler, 2022; Stallen et al., 2018).

Moreover, the psychological mechanisms underlying second- and third-party punishment seem to be quite different. Victims tend to focus primarily on the harm caused (Bellucci et al., 2020), and, thus, the consequences of the offense (Rim et al., 2013), ultimately affecting their motivation to punish (Yang et al., 2024). Specifically, second-party punishment seems to be more strongly motivated by antisocial or retributive motives than third-party punishment (Carpenter & Matthews, 2012; Jensen, 2010; Zhou et al., 2017). Additionally, the experience of harm and suffering triggers anger as the primary emotion underlying second-party punishment (Hartsough et al., 2020; Hopfensitz & Reuben, 2009; Russell & Giner-Sorolla, 2013; Seip et al., 2014). Based on the idea that identifiability intensifies the prevailing emotions and behavioral responses (Lewinsohn-Zamir et al., 2017;

Small & Loewenstein, 2005) offender identifiability should lead to stronger punishment (and more negative perception) by victims.

For uninvolved third parties, by contrast, punishment is relatively more strongly influenced by information about the offender's intentions (Bellucci et al., 2020; Yang et al., 2024). Third-party punishers, thus, tend to focus more on the cause of an act than on its consequences (Rim et al., 2013). This may be affected by the identification of the offender, as identifiability increases social and psychological closeness and makes the offending person ~~that has committed the offense~~ more concrete (Mentovich et al., 2016; Pałka et al., 2023), ultimately changing people's perceptions and attitudes toward them (Ledgerwood et al., 2010). Indeed, research shows that third parties show more benevolent attitudes toward socially and psychologically close people and punish them less strongly (Passarelli & Buchanan, 2020; Qu et al., 2018). Likewise, qualitative research shows that, as people gain more information about offenders, their attention moves away from the offense itself. Instead, they develop an understanding of the offender's reasons for committing the crime, leading to increased empathy toward them (Mae Boag & Wilson, 2013). Literature further suggests a complex interplay of emotions underlying third-party punishment, with moral outrage[2] caused by the perception of the offense (Camerer, 2003; Ginther et al., 2022; Hartsough et al., 2020; Landmann & Hess, 2017), but also empathy-related emotions such as sympathy, compassion, or pity for the offender playing a pivotal role (Kogut, 2011b; Rudolph et al., 2004). Importantly, while moral outrage is positively related to punishment (Carlsmith et al., 2002; Lotz et al., 2011), offender empathy is negatively related to punishment (Condon & DeSteno, 2011; Klimecki et al., 2016; Kogut, 2011b). Consequently, offender identifiability should

[2] Moral outrage is sometimes equated with anger and their distinction is sometimes questioned (Batson et al., 2007; O'Mara et al., 2011). Nevertheless, their distinctiveness has been emphasized in studies that differentiate between the two emotions (Ginther et al., 2022; Hechler & Kessler, 2018). Moral outrage, which encompasses the emotion of anger, arises from the perceived wrongfulness and immorality of an action, unlike anger which is typically triggered by its consequences (Hechler & Kessler, 2018).

rather lead to weaker punishment (and a more positive perception) of the offender by initially uninvolved third parties.

Importantly, the mixed results of research on the effects of offender identifiability, which was outlined above, are very much in line with this reasoning. Studies examining victim's responses to identified (vs. unidentified) offenders show negative consequences of identifiability for the offender (e.g., stronger punishment) (Kogut, 2011b; Small & Loewenstein, 2005). Whereas studies examining third- -parties' responses to identified (vs. unidentified) offenders show positive consequences of identifiability for the offender (e.g., weaker punishment) (Barak-Corren & Lewinsohn-Zamir, 2019; Baron & Ritov, 2009; Lewinsohn-Zamir et al., 2017). The present research builds upon this tentative evidence and, thus, is designed to integrate the (mixed) findings from the literature by directly testing the moderating influence of the role of the punisher in the initial offense on the effects of offender identifiability on punishment. Specifically, we hypothesize that the influence of offender identifiability on punishment intentions depends on the role of the punisher in the initial offense (Hypothesis 1). Victims should indicate stronger intentions to punish an identified compared to an unidentified offender. This effect should be reversed for , whereas third parties, who are expected to should indicate weaker intentions to punish an identified compared to an unidentified offender.

Notably, research by Kogut (2011b) is closely related to the present study. In this research, the author tested to what extent the effect of offender identifiability on punishment intentions depends on the perspective taken in an offense. More precisely, participants were asked to either take the perspective of the *victim* or the *offender* of a misconduct that was committed by an identified vs. unidentified offender (Kogut, 2011b, Study 1). Results revealed that participants had stronger intentions to punish identified offenders, compared to unidentified offenders, when adopting the victim-'s perspective, whereas they had weaker

intentions to punish identified offenders, compared to unidentified offenders, when taking the offender'’s perspective. In a subsequent study, the role of emotions in the decision to punish identified (vs. unidentified) offenders was investigated, without having participants adopt a specific perspective (Kogut, 2011b, Study 2). Therefore, in this study, participants may be considered third parties. From this third-party perspective, participants showed weaker intentions to punish identified offenders than unidentified ones. Again, this research may be interpreted as further tentative evidence in line with the present hypothesis. However, it did not directly compare the perspective of the victim with the perspective of an uninvolved third party and can, thus, not directly integrate the mixed literature by testing the hypothesis outlined above. This is the goal of the present study.

Psychological Mechanisms

Additionally, the present research seeks to enhance our understanding of the psychological mechanisms underlying the effects of offender identifiability. Affective mechanisms have been frequently proposed as a key aspect in this regard (Erlandsson et al., 2015; Lee & Feeley, 2016; Lewinsohn-Zamir et al., 2017; Loewenstein et al., 2005). One relevant affective response often discussed and found to be intensified by identification is empathy (Kogut & Ritov, 2005a; Moche et al., 2024; Sah & Loewenstein, 2012), which can influence both helping (Batson et al., 1981, 2001; Oceja et al., 2014) and punishing behavior (Lu & McKeown, 2018; Vitaglione & Barnett, 2003). For punishment, empathy towards the offender, that is, individuals’ capacity to understand an offender’s perspective and emotions, is particularly important (Kogut, 2011b; Unnever & Cullen, 2009). In the context of offender identifiability, research ~~has already shown~~suggests that victims express less empathy toward identified vs. unidentified offenders and, consequently, indicate stronger punitive intentions (Kogut, 2011b). Empathy for the offender was, thus, ~~found~~ suggested as a psychological mechanism (i.e., a mediator) between identification and punishment intentions. By contrast,

for uninvolved third parties, identifying information about the offender increased positive emotions toward the offender, including empathy (Barak-Corren & Lewinsohn-Zamir, 2019). Building on these findings and our theoretical considerations, we hypothesize that the influence of offender identifiability on empathy toward the offender depends on the role of the evaluating individual in the initial offense (Hypothesis 2). Victims should express less empathy toward an identified compared to an unidentified offender. This effect should be reversed for ~~whereas~~ third parties, who are expected to ~~should~~ express more empathy toward an identified compared to an unidentified offender. ~~Additionally, we hypothesize that offender empathy mediates the influence of offender identifiability on punitive intentions depending on the role of the punisher (Hypothesis 5a). Specifically, we assume that victims express less empathy toward an identified compared to an unidentified offender, whereas third parties express more empathy toward an identified compared to an unidentified offender. Offender empathy is then negatively related to punishment intentions.~~

Moral outrage, as another important affective response to an offense, arises from the perception that a fundamental moral principle has been violated, leading to negative emotions like indignation and anger (Batson et al., 2007). Anger, as a close sibling construct to and one part of moral outrage, has already been studied in the context of offender identifiability: Small and Loewensteins' study (2005), reported that anger served as a mediator between offender identifiability and second-party punishment. Specifically, victims felt more anger toward identified compared to unidentified offenders, and, in turn, punished them more strongly. Conversely, research on third-party evaluations ~~has demonstrated~~ indicates the direct opposite pattern, ~~showing that~~ with people expressing less anger and willingness to punish identified compared to unidentified offenders (Baron & Ritov, 2009, Study 8; Kogut, 2011b, Study 2). Importantly, even though the role of moral outrage has not yet been studied in the context of identifiability effects, it has been found to play an important mediating role in third-party punishment decisions (Bastian et al., 2013; Carlsmith et al., 2002; Ginther et al., 2022;

Landmann & Hess, 2017; Pfattheicher et al., 2019). In the present study, we examine moral outrage as a mediator underlying the effect of offender identifiability and the role of the punisher in the initial offense on punishment intentions. Accordingly, we hypothesize that the influence of offender identifiability on moral outrage toward the offender depends on the role of the evaluating individual in the initial offense (Hypothesis 3). Victims should express more moral outrage toward an identified compared to an unidentified offender. This effect should be reversed for ~~, whereas~~ third parties, who are expected to ~~should~~ express less moral outrage toward an identified compared to an unidentified offender. ~~Additionally, we hypothesize that moral outrage mediates the influence of offender identifiability on punitive intentions depending on the role of the punisher (Hypothesis 5b). Specifically, we assume that victims express more, whereas third parties express less moral outrage toward an identified compared to an unidentified offender. Moral outrage is then positively related to punishment intentions.~~

Another important variable for punishment decisions in general that will be considered in the present research, ~~which we also aim to consider in this research,~~ is offender blaming. Blame is typically assigned in response to actions or behaviors perceived as immoral (Squires, 1968). The attribution of blame plays a major role in punishment decisions (Sifferd, 2024), as it directs moral responsibility to the individual showing the behavior (i.e., the offender; Gray et al., 2012; Schein & Gray, 2014), ultimately increasing people's punitive intentions. This applies to both victims and third parties (Fehr & Fischbacher, 2004; Moore, 2020) and depends on the amount of harm caused, as well as the perceived intentionality of the offender's behavior (Krueger & Hoffman, 2016). ~~-~~Given that victims of a crime tend to focus more on the harm caused when assigning punishment, whereas third parties tend to focus more on the intentional agent (Bellucci et al., 2020), we suggest that blaming an identified vs. unidentified offender differs depending on whether the situation is evaluated by a victim or a third party. Barak-Corren and Lewinsohn-Zamirs (2019) already found that uninvolved third parties blamed an identified offender less compared to an unidentified offender. In sum, we

hypothesize that the influence of offender identifiability on offender blaming depends on the role of the evaluating individual in the initial offense (Hypothesis 4). Victims should blame an identified offender more compared to an unidentified offender. This effect should be reversed for ~~, whereas~~ third parties, who are expected to ~~should~~ blame an identified offender less compared to an unidentified offender. ~~Additionally, we hypothesize that offender blaming mediates the influence of offender identifiability on punitive intentions depending on the role of the punisher (Hypothesis 5c). Specifically, we assume that victims blame an identified compared to an unidentified offender more, whereas third parties blame an identified compared to an unidentified offender less. Offender blaming is then positively related to punishment intentions.~~

The Present Research

In summary, the present research aims to enhance the understanding of the identifiability effect, specifically focusing on offender identifiability, thereby extending a comparatively limited literature, and integrating mixed results that were reported so far. We examine the role of both offender identifiability (unidentified vs. ~~un~~identified) and the role of the punisher (victim vs. third -party) on punishment intentions. ~~Moreover, we investigate the mediating role of relevant emotions like empathy, moral outrage, and blame in this process.~~ In total, we will test the following hypotheses, which are visually illustrated in Figure 1: ~~In total, we will test the following hypotheses:~~

H1: The influence of offender identifiability on punishment intentions depends on the role of the punisher in the initial offense. Victims indicate stronger intentions to punish an identified compared to an unidentified offender. ~~,~~This effect should be reversed for ~~whereas~~ third parties, who are expected to indicate weaker intentions to punish an identified compared to an unidentified offender.

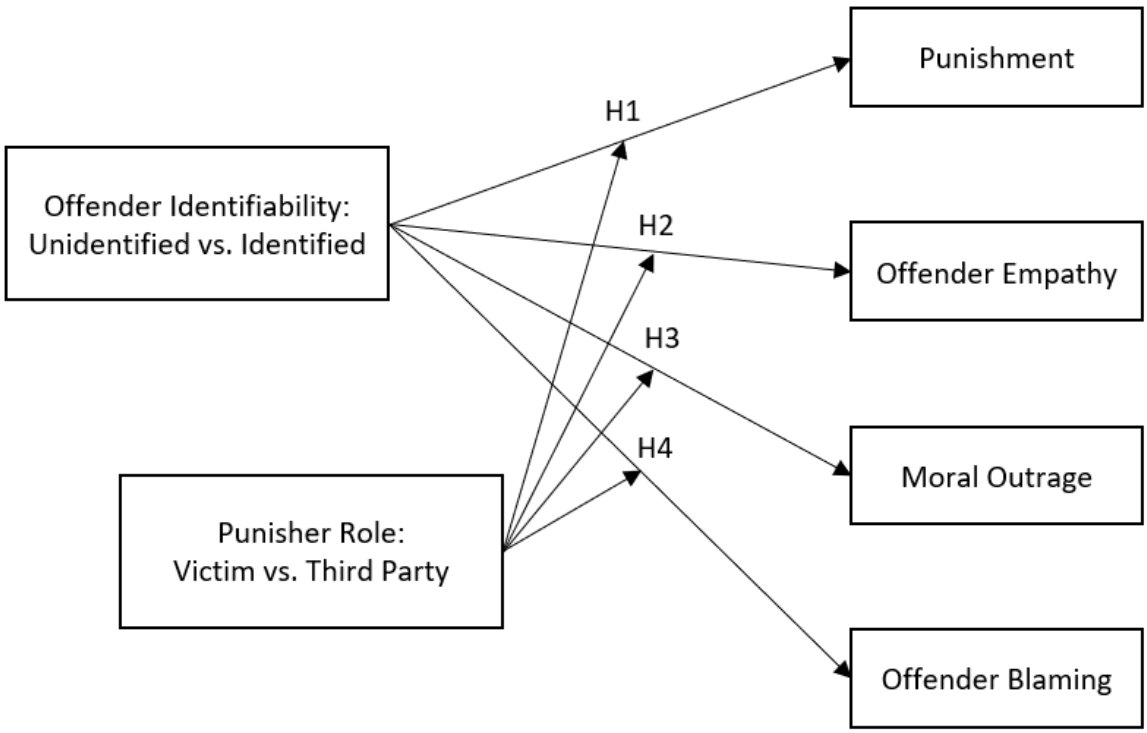
H2: The influence of offender identifiability on empathy toward the offender depends on the role of the evaluating individual in the initial offense. Victims express less empathy toward an identified compared to an unidentified offender. This effect should be reversed for ~~,~~ ~~whereas~~ third parties, who are expected to express more empathy toward an identified compared to an unidentified offender.

H3: The influence of offender identifiability on moral outrage toward the offender depends on the role of the evaluating individual in the initial offense. Victims express more moral outrage toward an identified compared to an unidentified offender. This effect should be reversed for~~,~~ ~~whereas~~ third parties, who are expected to express less moral outrage toward an identified compared to an unidentified offender.

H4: The influence of offender identifiability on offender blaming depends on the role of the evaluating individual in the initial offense. Victims blame an identified offender more compared to an unidentified offender. This effect should be reversed for~~,~~ ~~whereas~~ third parties, who are expected to blame an identified offender less compared to an unidentified offender.

Figure 1

Conceptual model of the hypothesized interaction effects of punisher role and offender identifiability on punishment intentions, offender empathy, moral outrage, and offender blaming.



~~**H5:** (a) Offender empathy, (b) moral outrage, and (c) offender blaming are mediators for the effect of the interaction of identifiability and initial role of the punisher on punishment. Victims express less empathy, more moral outrage, and more offender blaming to an identified compared to an unidentified offender, whereas third parties express more empathy, less moral outrage, and less offender blaming to an identified compared to an unidentified offender. Offender empathy is then negatively related to punishment intentions, whereas moral outrage and offender blaming are positively related to punishment intentions.~~

Methods and Materials

We created a simulated dataset using the statistics program R (simulation code available on Open Science Framework[3]). Using this dataset, we conducted all planned calculations and report the results below as an example of how they will be conducted and reported after the

[3] https://osf.io/h5uys/?view_only=60348d468d8949aab31e9304128ec76a

actual data collection. Consequently, the following sections are written in past tense, even though no actual data has been collected yet.

To investigate the role of offender identifiability on punishment intentions of both affected victims and uninvolved third parties, we ~~plan to~~ conducted an online study with a crime scenario. In this scenario, we ~~will~~ experimentally manipulated whether the offender was~~is~~ identified (vs. not) and whether participants ~~are~~ were victims or third parties. Hence, we ~~will have~~had a 2 (offender identifiability: no vs. yes) x 2 (punisher role: victim vs. third party) between-subjects design. Participants ~~will~~ were~~be~~ randomly assigned to one of these experimental conditions. All materials, data, and analyses scripts ~~will be provided~~can be found in the supplementary documents on the Open Science Framework (OSF)[4].

Sample, Data Collection, and Exclusion Criteria

We conducted an a priori power analysis to determine the required sample size to test our central hypothesis (Hypothesis 1) using the R package "pwr"~~G*Power~~ (Champely, 2020). To ensure sufficient power for our study, we based our analysis on the effects of offender identifiability on punishment reported in the literature, ranging from small (Barak-Corren & Lewinsohn-Zamir, 2019, $f = 0.12$; Lewinsohn-Zamir et al., 2017, $f = 0.10$), to medium~~-~~-sized effects (Small & Loewenstein, 2005, $f = 0.26$; Kogut, 2011b, Study 2, $f = 0.32$). However, given that this prior research only focused on either the victims or third parties as punishers, this only corresponds to the planned follow-up analyses (see below), but not our planned ANOVA testing the interaction effect of offender identifiability and role of the punisher on punishment intentions. Consequently, we planned to collect enough data to detect a small effect ($f = 0.125$) in a 2 x 2 between-subjects-design ANOVA, with a power of $1-\beta = .90$ and $\alpha = .05$. An a priori power analysis resulted in a required sample size of N = 911~~675~~ participants. To account for dropouts or data exclusions based on inattentive participation (see

[4] https://osf.io/h5uys/?view_only=60348d468d8949aab31e9304128ec76a

below for more information on ~~planned~~ data exclusions), we aim~~ed~~ to recruit N = ~~750~~ 1000 participants[5] finishing the study up until the last relevant item (so-called "use-me" item, see below). The code to replicate this a priori power analysis can be found in our analysis script on the OSF.

We recruited participants from Prolific Academic (Palan & Schitter, 2018). Participants had to be at least 18 years of age and fluent in English to take part in the study. Initially, we recruited N = 1000 participants with complete data, of which N = 62 were excluded because of the following pre-registered criteria: (1) completing the survey too fast[6] (n = 22), (2) answering one of three attention checks incorrectly (n = 37), and (3) indicating their data should not be used (Meade & Craig, 2012) (n = 3). This resulted in a final sample size of N = 938 participants (46.27% female, 53.20% male, 0.53 % other; age range: 18 - 68; M = 35.05, SD = 9.31).

The study ~~will be~~was conducted online via SoSciSurvey[7] (Leiner, 2024)~~. We plan to recruit participants from Prolific Academic. The planned study will be conducted~~ and in full accordance with the Ethical Guidelines of the American Psychological Association (APA) and the German Association of Psychologists (DGPs). Ethics approval by an institutional review board or committee is not mandatory at German universities if a study fully discloses all information regarding the study to participants and is unlikely to cause harm, stress, or any other form of negative affect exceeding an "everyday experience" level. This ~~will not be~~was not the case for the present study; therefore, no official ethics approval needed~~s~~ to be obtained. Additionally, participants ~~will not be~~were not deceived in this study.

[5] We do this based on own experiences with previous studies and literature, suggesting to account for dropout and exclusion rates of around 10% (Hoerger, 2010; Meade & Craig, 2012).
[6] The online platform SosciSurvey provides an index of relative completion speed. This index is calculated based on median completion times across the survey. An index > 2 means that, across all pages of the survey, a participant completed the pages at least twice as fast as the typical participant of this survey (Leiner, 2019).
[7] A pretest version of the questionnaire can be found via the following link: https://survey.ifkw.lmu.de/JudgmentRevealed/?act=1Zng3r85AfE2K8wgGWd1wSEc

~~Participants must be at least 18 years of age and fluent in English to take part in the study. Additionally, we will exclude participants who either (a) complete the survey too slowly or too fast (b) answer one of three attention check questions incorrectly, or (c) indicate that they did not participate conscientiously and that their data should not be used at the end of the questionnaire.~~

Design and Measures

All materials of the ~~planned~~ study can be found in our supplementary materials on the OSF[8]. The experiment ~~will~~ followed a 2 x 2 between-subjects design with offender identifiability (no vs. yes) and punisher role (victim vs. third party) as independent variables and punishment intentions, offender empathy, moral outrage, and offender blaming as measured variables. After providing informed consent and demographic information (age and gender), participants ~~will~~ read a scenario detailing a pickpocketing offense. We chose pickpocketing because this typically does not involve direct confrontation between the victim and the offender in the situation of the offense (as compared to, for example, an assault). Therefore, this scenario allows the offender to be presented as either identified or unidentified for both the victim and third-party role. In this scenario, participants ~~will~~ were either ~~be~~ in the role of the victim of the pickpocket or the role of an uninvolved third party. Next, participants ~~will~~ indicated their intentions to punish the offender, before answering questions on offender empathy, moral outrage, and offender blaming in randomized order. Following this, we ~~will~~ assessed participants' perceived seriousness of the behavior, harm caused, intentionality of the offender, and their subjective psychological closeness toward the offender for exploratory purposes, again in randomized order. Following three attention checks on the content of the scenario, participants ~~will~~ indicated whether they participated conscientiously (so-called "use-me" item; Meade & Craig, 2012), before they ~~will be~~were fully debriefed and thanked.

[8] https://osf.io/h5uys/?view_only=60348d468d8949aab31e9304128ec76a

Manipulation

We ~~will~~ used self-created scenarios to experimentally manipulate whether the offender was~~is~~ unidentified (n = 481) or ~~un~~identified (n = 457) and whether participants ~~are~~ were victims (n = 483) or third parties (n = 455). In the scenario, participants read about a pickpocketing incident, in which the victim had placed their backpack next to them while locking their bike. When turning back to the backpack again after locking the bike, they realized that their backpack was missing, with a person running away with it. The offender was~~is~~ caught shortly afterward by the police. In the offender unidentified condition, offenders ~~will~~ were only ~~be~~ termed as "the offender". In the offender identified condition, offenders ~~will be~~were identified by name, age, and a picture[9]. Gender of the offender was counterbalanced. Names and ages were randomly selected from six possible variations each and we used race-neutral names taken from previous research (Darolia et al., 2016)~~.~~. Pictures were ~~(~~taken from the ~~Basel~~ Chicago Face Database~~;~~ (Ma et al., 2015). To address potential race effects, we randomly selected three female and three male photos from each of the following ethnic groups: Asian, Black, Latino, and White. This selection includes individuals aged between 24 and 29. Consequently, our sample of pictures consists of 12 female and 12 male pictures. One of these was randomly presented to the participants. ~~Names, ages, and pictures will be randomly selected from six possible variations each. Gender of the offender will be counterbalanced.~~

Measured variables

Unless stated otherwise, all dependent variables ~~will be~~were measured on a 6-point Likert scale ranging from 1 = "Strongly Disagree" to 6 = "Strongly Agree". Items within

[9] We provide name, age, and a picture as these elements are often reported in media coverage on offenses and provided in research on the identifiability effect (see e.g., Genevsky et al., 2013; Kogut, 2011a; Wiss et al., 2015).

scales were ~~will be~~ presented in randomized order. All scales are based on prior research to decrease the chances of potential floor or ceiling effects.

We ~~will~~ measured participant~~'s~~' intentions to punish the offender with three items from Pfattheicher et al. (2019). The items include "The offender deserves ~~hard~~ harsh treatment.", "The offender should be taught a lesson.", and "The offender should be severely punished" ($\alpha = 0.02$, $\omega = 0.36$). Empathy toward the offender ~~will be~~was measured with six items that were adapted from Haegerich and Bottoms (2000) and Pfattheicher et al., (2019) and extend. These items read as follows: "I feel compassion for the offender.", "I can experience the same feelings as the offender experienced", "I feel like I can easily take the perspective of the offender", "I can imagine the thoughts going through the offender's mind"; "I emphasize with the offender.", and "I genuinely understand the offender" ($\alpha = 0.06$, $\omega = 0.11$). This measure of empathy captures both the affective and cognitive components of empathy (Cuff et al., 2016; Davis et al., 1987; Duan & Hill, 1996). Moral outrage ~~will be~~was measured with three items from Pfattheicher et al. (2019), including "I am angry ~~at~~ with the offender", "I am outraged by the offender's behavior.", and "The offender's behavior was absolutely immoral." ($\alpha = 0.04$, $\omega = 0.37$). We ~~will~~ further measured offender blaming with two items adapted from Kogut (2011a): "The offender is responsible for what happened." and "The offender is to blame for what happened" ($\alpha = 0.02$, $\omega = 0.01$).

Several exploratory variables ~~will be~~were assessed as follows: Perceived seriousness of the behavior ~~will be~~was measured with a single item adapted from Herzog and Oreg (2008): "How serious was the described behavior?", on a scale ranging from 1 = "Not at all serious" to 6 = "Very serious". Perceived harm ~~will be~~was measured with one item adapted from Bastian et al. (2013) asking "How much harm has been caused as a result of the behavior?" ranging from 1 = "No harm at all" to 6 = "A lot of harm". Additionally, participants' perceived intentionality of the offender's behavior ~~will be~~was measured with a

single item from Struthers et al. (2008) asking "To what extent do you think the offender intended to do what he/she did?" ranging from 1 = "Not at all intended" to 6 = "Completely intended". Participants' perceived psychological closeness toward the offender ~~will be~~was assessed with three items, adapted from Yan et al. (2016). These items include "I feel similar to the offender.", "I feel like I know the offender well.", and "I feel close to the offender", and ~~will be~~was answered on a scale ranging from 1 = "Strongly Disagree" to 6 = "Strongly Agree" (α = -0.12, ω = 0.37). The attention checks ~~will be~~were assessed with three single choice items, asking (1) what was stolen in the scenario, (2) what role the participant had, and (3) what kind of information was given about the offender in the described scenario. We provided three answer options for each attention check item, with one answer that is clearly correct.

Statistical Analysis

All statistical analyses ~~will be~~were performed using "R" version 4.4.0 and a *p*-value of .05 or less ~~will be~~was considered to be statistically significant. ~~We will report the demographic information (age and gender), as well as the number of participants per experimental condition in our sample.~~ Scale scores for all self-report measures ~~will be~~were calculated by averaging items. ~~and we will report internal consistency (McDonald's ω) of all scales in the method section. In addition, we provide mean values and standard deviations of all measured variables, as well as correlations between the variables for the overall sample. Similarly, we will report mean values and standard deviations of all measured variables separated for conditions.~~

Table 1 shows means, standard deviations, and Pearson correlations between the measured variables. Notably, none of these correlations was significant, except for the correlation between offender blaming and moral outrage (r = -.07). This suggests that as

offender blaming increases, moral outrage decreases slightly, although this relationship is relatively weak.

Table 1

Means, standard deviations, and correlations with confidence intervals.

Var	*M*	*SD*	1	2	3	4	5	6	7
1. PI	3.44	0.99							
2. OE	3.51	0.72	-.02 [-.09, .04]						
3. MO	3.46	1.00	.04 [-.03, .10]	.01 [-.05, .08]					
4. OB	3.48	1.23	-.01 [-.07, .06]	-.07* [-.13, -.00]	.05 [-.02, .11]				
5. SE	3.56	1.71	.00 [-.06, .06]	.02 [-.04, .09]	.03 [-.04, .09]	-.00 [-.06, .06]			
6. HA	3.58	1.72	.01 [-.05, .08]	-.02 [-.09, .04]	-.01 [-.07, .06]	.04 [-.03, .10]	-.01 [-.07, .06]		
7. IN	3.41	1.71	-.02 [-.08, .04]	-.04 [-.11, .02]	.00 [-.06, .07]	.02 [-.04, .08]	-.05 [-.12, .01]	-.04 [-.10, .03]	
8. PC	3.49	0.94	-.06 [-.13, .00]	.00 [-.06, .06]	-.04 [-.11, .02]	.03 [-.04, .09]	-.03 [-.10, .03]	-.02 [-.09, .04]	.03 [-.03, .10]

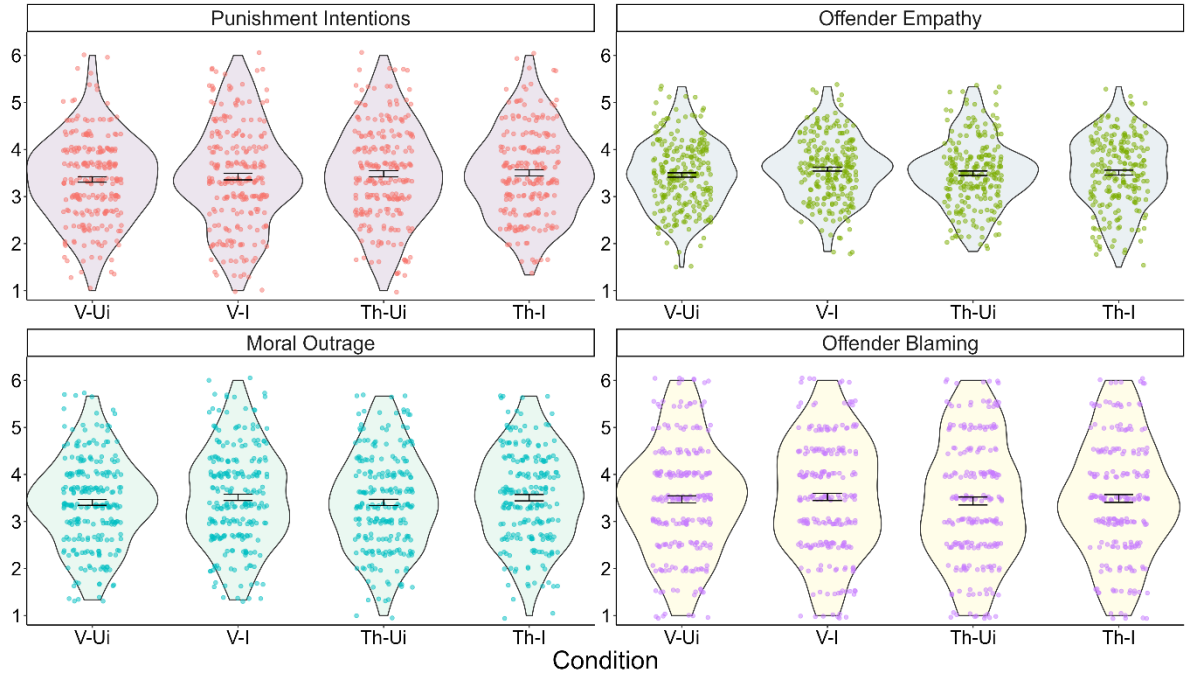
Note. N = 938; M and SD are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. PI = Punishment Intentions, OE = Offender Empathy, MO = Moral Outrage, OB = Offender Blaming, SE = Seriousness, HA = Harm, IN = Intentionality, PC = Psychological Closeness. * indicates $p < .05$. ** indicates $p < .01$.

Figure 2 displays means and distributions of punishment intentions, offender empathy, moral outrage, and offender blaming for each condition. Before conducting any hypotheses tests, we tested the assumption of homogeneity of variances using Levene's test for all dependent variables. Based on these tests, the assumption of homogeneity of variances was

confirmed for punishment intentions (p = .072), offender empathy (p = .193), moral outrage (p = .206), and offender blaming (p = .212). Therefore, we conducted conventional Fisher's ANOVAs to test our hypotheses for all these variables[10]. In the following, we report our findings for each dependent variable separately.

Figure 2

Means and distributions of punishment intentions, offender empathy, moral outrage, and offender blaming for each condition. Error bars represent one standard error of the mean.



Note. V-Ui = Victim role, Unidentified offender; V-I = Victim role, Identified offender; Th-Ui = Third Party role, Unidentified offender; Th-I = Third Party role, Identified offender.

Punishment intentions

~~We will~~As depicted in Figure 2 (upper left corner), punishment intentions were similar in all conditions. To test our central prediction (Hypothesis 1:~~;~~ T~~t~~he effect of offender identifiability on peoples' punishment intentions depends on the role of the punisher in the initial offense), ~~by~~we conducted~~ing~~ a two-way ANOVA with both experimental factors and

[10] If Levene's tests indicate a violation of the homogeneity of variances assumption, we will perform Robust ANOVAs using the WRS2 package (Mair & Wilcox, 2020) and report the results of both the conventional Fisher's and robust ANOVA.

their interaction as predictors. ~~including interaction.~~ We did not find a significant main effect of role, $F(1,934) = 2.34$, $p = .127$, $\eta^2_p < 0.01$, 95% CI [0.00;1.00], or offender identifiability on punishment intentions, $F(1,934) = 0.37$, $p = .543$, $\eta^2_p < 0.01$, 95% CI [0.00;1.00]. Importantly, and contrary to our central prediction, the interaction effect of role and identifiability on punishment intentions was not significant, $F(1,934) = 0.11$, $p = .746$, $\eta^2_p < 0.01$, 95% CI [0.00;1.00].[11] ~~The interaction effect is most central for our hypothesis. If the interaction effect (or any of the main effects) is significant, we will conduct follow-up independent samples *t*-tests (one-sided) or Welch's t-tests (in case of variance heterogeneity) to determine the direction of the effect. If we find a significant interaction effect and differences in the hypothesized direction, we take this as an indication that the effect of offender identifiability on punishment intentions depends on the role of the punisher and that victims indicate stronger intentions to punish identified (vs. unidentified) offenders, while third parties indicate weaker intentions to punish identified (vs. unidentified) offenders.~~

Psychological Mechanisms

To further ~~investigate~~ test our hypotheses about the interaction between offender identifiability and the punisher's role on relevant affective and cognitive mechanisms (Hypotheses 2-4), we ~~will~~ conducted a series of two-way ANOVAs. For offender empathy, conditions did not differ descriptively, as depicted in Figure 2 (upper right corner). In line with this pattern, we did not find a significant main effect of role, $F(1,934) = 0.15$, $p = .699$, $\eta^2_p < 0.01$, 95% CI [0.00;1.00] or offender identifiability on offender empathy, $F(1,934) = 2.39$, $p = .123$, $\eta^2_p < 0.01$, 95% CI [0.00;1.00]. Additionally, the predicted interaction effect of

[11] Note that if the interaction effects for Hypotheses 1-4 are significant, we will conduct follow-up one-sided Welch's *t*-tests for independent samples to test whether the directions of the effects are as hypothesized. If we find significant main effects, we will perform (additional) pairwise Welch's *t*-tests to explore specific group differences. In all cases, we will report the effect sizes and confidence intervals of these analyses.

role and identifiability on offender empathy was also not significant, $F(1,934) = 1.42$, $p = .234$, $\eta^2_p < 0.01$, 95% CI [0.00;1.00].

As depicted in Figure 2 (lower left corner), conditions did not differ descriptively in expressed moral outrage. Again, our analyses did not reveal any significant main effects of role, $F(1,934) = 0.00$, $p = .951$, $\eta^2_p < 0.01$, 95% CI [0.00;1.00] or offender identifiability on moral outrage, $F(1,934) = 2.55$, $p = .111$, $\eta^2_p < 0.01$, 95% CI [0.00;1.00]. There was also no significant interaction effect, $F(1,934) = 0.00$, $p = .948$, $\eta^2_p < 0.01$, 95% CI [0.00;1.00].

Finally, conditions also did not differ descriptively regarding offender blaming (Figure 2, lower right corner). In line with this, there was no significant main effect of role on offender blaming, $F(1,934) = 0.21$, $p = .651$, $\eta^2_p < 0.01$, 95% CI [0.00;1.00]. Similarly, the main effect of offender identifiability was not significant, $F(1,934) = 0.43$, $p = .511$, $\eta^2_p < 0.01$, 95% CI [0.00;1.00]. Furthermore, and again contrary to our prediction, the interaction effect of role and identifiability on offender blaming was not significant, $F(1,934) = 0.00$, $p = .987$, $\eta^2_p < 0.01$, 95% CI [0.00;1.00]. ~~For each variable (offender empathy, moral outrage, offender blaming), again, we will conduct follow-up independent samples *t*-tests (one-sided) or Welch's t-test (if variance homogeneity is not given), to determine whether the direction of the effect is in line with our hypotheses.~~

~~To examine the potentially mediating role of offender empathy, moral outrage, and offender blaming in the relationship between offender identifiability and the punisher's role on punishment intentions (Hypothesis 5 a-c), we will perform one bootstrapped moderated mediation analysis for detecting indirect effects using the lavaan package . Offender identifiability will be treated as independent variable; punisher role will be treated as moderator variable; offender empathy, moral outrage, and offender blaming will be treated as parallel mediator variables; punishment intentions will be treated as the dependent variable. All continuous variables will be standardized. When 95% confidence intervals for the respective indirect effects do not contain zero (and estimates are in the hypothesized direction), we take this as an indication that the influence of offender identifiability and punisher's role on punishment intentions is mediated by (a) offender empathy, (b) moral outrage, and (c) offender blaming, respectively.~~

Contributor Roles Taxonomy

Role	Theresa Blanke	Mathias Twardawski
Conceptualization	X	X
Data curation	X	
Formal analysis	X	
Funding acquisition		X
Investigation	X	
Pre-registration peer review / verification	X	X
Data analysis peer review / verification	X	X
Methodology	X	X
Project administration	X	X
Resources	X	
Software	X	
Supervision		X
Validation		X
Visualization	X	
Writing-original draft	X	
Writing-review and editing		X

Study Design Template

Question	Hypothesis	Sampling plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes
How does identifying an offender affect victims' and third~~-~~parties' punishment intentions?	**H1:** The influence of offender identifiability on punishment intentions depends on the role of the punisher in the initial offense. Victims indicate stronger intentions to punish an identified compared to an unidentified offender~~.~~, This effect should be reversed for ~~whereas~~ third parties, who are expected to indicate weaker intentions to punish an identified compared to an unidentified offender.	We aim to recruit 1000 participants via Prolific Academic ~~plan to collect enough data~~ to detect a small effect (f = 0.125) in a 2 x 2 between-subjects-design ANOVA, with a power of 1-β = .90 and α = .05. The~~An~~ a priori power analysis resulted in a required sample size of N = 911~~675~~ participants. To account for dropouts or data exclusions, we increased the planned sample size to N = 1000 (see manuscript for further explanation). ~~based on inattentive participation (see manuscript for more information on planned data exclusions), we aim to recruit N = 750 participants. Participants will be recruited via Prolific Academic.~~	We will perform an ANOVA with identifiability (IV1), punisher role (IV2), and their interaction (IV3) on punishment intentions (DV). If we find an interaction effect, we will perform independent samples ~~t-tests or~~ Welch's t-tests. ~~(if variance homogeneity is not given).~~	To ensure sufficient power to test the central prediction (H1), ~~of our study,~~ we based our power analysis on the effects of offender identifiability on punishment reported in the literature, ranging from small (Barak-Corren & Lewinsohn-Zamir, 2019, f = 0.12; Lewinsohn-Zamir et al., 2017, f = 0.10), to medium sized effects (Small & Loewenstein, 2005, f = 0.26; Kogut, 2011b, Study 2, f = 0.32). However, ~~given that this~~these prior ~~research only~~studies focused on~~ly on either the~~ victims or third parties as punishers, ~~this only corresponds to the planned follow-up analyses, but not our planned ANOVA testing the~~not the interaction effect of offender identifiability and punisher role ~~of the punisher~~ on punishment intentions. Consequently, we plan to collect enough data to detect a small effect (f = 0.125) in a 2 x 2 between-subjects-design ANOVA, with a power of 1-ß = .90 and α = .05. ~~Since punishment intentions is our central dependent variable, all~~All	If results are in line with our predictions, this helps identifying an important boundary condition of the effect of offender identifiability on punishment: It depends on the role of the punisher in the initial offense (second vs. third parties). If results are not in line with our predictions, the role of the punisher in the initial offense does either not affect the effect of offender identifiability on punishment intentions, or offender identifiability does not at all affect people's punishment intentions.	Research on offender identifiability found mixed effects on punishment. The present study helps integrating this literature by identifying an important boundary condition: the role of the punisher in the initial offense. If we find the hypothesized interaction effect, this would show that the current idea that offender identifiability generally increases punishment intentions (Small & Loewenstein, 2005) is overly simple. If we do not find the hypothesized interaction effect, main effects of our manipulations may become more interesting and may replicate some results reported in the literature, while failing to replicate other results reported in the literature.
How does identifying an offender affect victims' and third parties' emotional and attitudinal responses towards them?	**H2:** The influence of offender identifiability on empathy toward the offender depends on the role of the evaluating individual in the initial offense. Victims express less empathy toward an identified compared to an unidentified offender~~.~~, This effect should be reversed for ~~whereas~~ third parties, who are expected to express more		We will perform an ANOVA with identifiability (IV1), punisher role (IV2), and their interaction (IV3) on offender empathy (DV). If we find an interaction effect, we will perform independent samples ~~t-tests or~~ Welch's t-tests ~~(if variance~~		If results are in line with our predictions, this helps identifying an important boundary condition of the effect of offender identifiability on evaluator's emotional and attitudinal responses: These responses depend on the role of the evaluator in the initial offense (second vs. third parties). If results are not in line with our predictions, the role of the	

	empathy toward an identified compared to an unidentified offender.		~~homogeneity is not given)~~.	other analys~~e~~is follow this reasoning.	evaluator in the initial offense does either not affect the effect of offender identifiability on offender empathy, moral outrage, or offender blaming, or offender identifiability does not at all affect these emotional and attitudinal responses.	
	H3: The influence of offender identifiability on moral outrage toward the offender depends on the role of the evaluating individual in the initial offense. Victims express more moral outrage toward an identified compared to an unidentified offender~~.~~, This effect should be reversed for ~~whereas~~ third parties, who are expected to ~~e~~express less moral outrage toward an identified compared to an unidentified offender.		We will perform an ANOVA with identifiability (IV1), punisher role (IV2), and their interaction (IV3) on moral outrage (DV). If we find an interaction effect, we will perform independent samples ~~t-tests or~~ Welch's *t*-tests ~~(if variance homogeneity is not given)~~.			
	H4: The influence of offender identifiability on offender blaming depends on the role of the evaluating individual in the initial offense. Victims blame an identified offender more compared to an unidentified offender~~.~~, This effect should be reversed for ~~whereas~~ third parties, who are expected to blam~~e~~e an identified offender less compared to an unidentified offender.		We will perform an ANOVA with identifiability (IV1), punisher role (IV2), and their interaction (IV3) on offender blaming (DV). If we find an interaction effect, we will perform independent samples ~~t-tests or~~ Welch's *t*-tests ~~(if variance homogeneity is not given)~~.			

References

Barak-Corren, N., & Lewinsohn-Zamir, D. (2019). What's in a name? The disparate effects of identifiability on offenders and victims of sexual harassment. *Journal of Empirical Legal Studies*, *16*(4), 955–1000. https://doi.org/10.1111/jels.12233

Baron, J., & Ritov, I. (2009). The role of probability of detection in judgments of punishment. *Journal of Legal Analysis*, *1*(2), 553–590. https://doi.org/10.1093/jla/1.2.553

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

Batson, C. D., Ahmad, N., Lishner, D. A., & Tsang, J.-A. (2001). Empathy and altruism. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 485–498). Oxford University Press. https://doi.org/10.1093/oso/9780195135336.003.0035

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation? *Journal of Personality and Social Psychology*, *40*(2), 290–302. https://doi.org/10.1037/0022-3514.40.2.290

Batson, C. D., Kennedy, C. L., Nord, L.-A., Stocks, E. L., Fleming, D. A., Marzette, C. M., Lishner, D. A., Hayes, R. E., Kolchinsky, L. M., & Zerger, T. (2007). Anger at unfairness: Is it moral outrage? *European Journal of Social Psychology*, *37*(6), 1272–1285. https://doi.org/10.1002/ejsp.434

Bellucci, G., Camilleri, J. A., Iyengar, V., Eickhoff, S. B., & Krueger, F. (2020). The emerging neuroscience of social punishment: Meta-analytic evidence. *Neuroscience & Biobehavioral Reviews*, *113*, 426–439. https://doi.org/10.1016/j.neubiorev.2020.04.011

Bystranowski, P., Janik, B., Próchnicki, M., Hannikainen, I. R., da Franca Couto Fernandes de Almeida, G., & Struchiner, N. (2022). Do formalist judges abide by their abstract

principles? A two-country study in adjudication. *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, *35*(5), 1903–1935. https://doi.org/10.1007/s11196-021-09846-6

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

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

Carpenter, J. P., & Matthews, P. H. (2012). Norm enforcement: Anger, indignation, or reciprocity? *Journal of the European Economic Association*, *10*(3), 555–572. https://doi.org/10.1111/J.1542-4774.2011.01059.x

Champely, S. (2020). *pwr: Basic functions for power analysis* (R-Paket Version 1.3-0) [Computer Software]. Comprehensive R Archive Network. https://cran.r-project.org/package=pwr

Civai, C., Huijsmans, I., & Sanfey, A. G. (2019). Neurocognitive mechanisms of reactions to second- and third-party justice violations. *Scientific Reports*, *9*, 9271. https://doi.org/10.1038/s41598-019-45725-8

Condon, P., & DeSteno, D. (2011). Compassion for one reduces punishment for another. *Journal of Experimental Social Psychology*, *47*(3), 698–701. https://doi.org/10.1016/j.jesp.2010.11.016

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review*, *8*(2), 144–153. https://doi.org/10.1177/1754073914558466

Darolia, R., Koedel, C., Martorell, P., Wilson, K., & Perez-Arce, F. (2016). Race and gender effects on employer interest in job applicants: new evidence from a resume field

experiment. *Applied Economics Letters*, *23*(12), 853–856. https://doi.org/10.1080/13504851.2015.1114571

Davis, M. H., Hull, J. G., Young, R. D., & Warren, G. G. (1987). Emotional reactions to dramatic film stimuli: The influence of cognitive and emotional empathy. *Journal of Personality and Social Psychology*, *52*(1), 126–133. https://doi.org/10.1037/0022-3514.52.1.126

Duan, C., & Hill, C. E. (1996). The current state of empathy research. *Journal of Counseling Psychology*, *43*(3), 261–274. https://doi.org/10.1037/0022-0167.43.3.261

Erlandsson, A., Björklund, F., & Bäckström, M. (2015). Emotional reactions, perceived impact and perceived responsibility mediate the identifiable victim effect, proportion dominance effect and in-group effect respectively. *Organizational Behavior and Human Decision Processes*, *127*, 1–14. https://doi.org/10.1016/J.obhdp.2014.11.003

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

Feng, C., Tian, X., & Luo, Y.-J. (2023). Neurocomputational substrates underlying the effect of identifiability on third-party punishment. *Journal of Neuroscience*, *43*(47), 8018–8031. https://doi.org/10.1523/jneurosci.0460-23.2023

Genevsky, A., Västfjäll, D., Slovic, P., & Knutson, B. (2013). Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving. *Journal of Neuroscience*, *33*(43), 17188–17196. https://doi.org/10.1523/jneurosci.2348-13.2013

Ginther, M. R., Hartsough, L. E. S., & Marois, R. (2022). Moral outrage drives the interaction of harm and culpable intent in third-party punishment decisions. *Emotion*, *22*(4), 795–804. https://doi.org/10.1037/emo0000950

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101–124. https://doi.org/10.1080/1047840X.2012.651387

Haegerich, T. M., & Bottoms, B. L. (2000). Empathy and jurors' decisions in patricide trials involving child sexual assault allegations. *Law and Human Behavior*, *24*(4), 421–448. https://doi.org/10.1023/A:1005592213294

Hartsough, L. E. S., Ginther, M. R., & Marois, R. (2020). Distinct affective responses to second- and third-party norm violations. *Acta Psychologica*, *205*, 103060. https://doi.org/10.1016/j.actpsy.2020.103060

Hechler, S., & Kessler, T. (2018). On the difference between moral outrage and empathic anger: Anger about wrongful deeds or harmful consequences. *Journal of Experimental Social Psychology*, *76*, 270–282. https://doi.org/10.1016/j.jesp.2018.03.005

Hechler, S., & Kessler, T. (2022). The importance of unfair intentions and outcome inequality for punishment by third parties and victims. *Zeitschrift Für Psychologie*, *230*(2), 114–126. https://doi.org/10.1027/2151-2604/a000458

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767–1770. https://doi.org/10.1126/science.1127333

Herzog, S., & Oreg, S. (2008). Chivalry and the moderating effect of ambivalent sexism: Individual differences in crime seriousness judgments. *Law & Society Review*, *42*(1), 45–74. https://doi.org/10.1111/j.1540-5893.2008.00334.x

Hoerger, M. (2010). Participant dropout as a function of survey length in internet-mediated university studies: Implications for study design and voluntary participation in

psychological research. *Cyberpsychology, Behavior and Social Networking*, *13*(6), 697. https://doi.org/10.1089/cyber.2009.0445

Hooffacker, G., & Meier, K. (2017). Publizistische Grundsätze (Pressekodex). In *La Roches Einführung in den praktischen Journalismus* (pp. 157–160). Springer VS. https://doi.org/10.1007/978-3-658-16658-8_8

Hopfensitz, A., & Reuben, E. (2009). The importance of emotions for the effectiveness of social punishment. *The Economic Journal*, *119*(540), 1534–1559. https://doi.org/10.1111/j.1468-0297.2009.02288.x

Hou, J., Zhang, C., Zhao, F., & Guo, H. (2023). Underlying Mechanism to the Identifiable Victim Effect in Collective Donation Action Intentions: Does Emotional Reactions and Perceived Responsibility Matter? *Voluntas*, *34*(3), 552–572. https://doi.org/10.1007/S11266-022-00480-9/TABLES/8

Hough, J. M., Jacobson, J., & Millie, A. (2003). *The decision to imprision: Sentencing and the prison population*. Prison Reform Trust.

Hsu, S.-L. (2008). The identifiability of bias in environmental law. *Florida State University Law Review*, *35*(2), 433–504.

Jensen, K. (2010). Punishment and spite, the dark side of cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1553), 2635–2650. https://doi.org/10.1098/rstb.2010.0146

Klimecki, O. M., Vuilleumier, P., & Sander, D. (2016). The impact of emotions and empathy-related traits on punishment behavior: Introduction and validation of the inequality game. *PLoS ONE*, *11*(3), e0151028. https://doi.org/10.1371/journal.pone.0151028

Kogut, T. (2011a). Someone to blame: When identifying a victim decreases helping. *Journal of Experimental Social Psychology*, *47*(4), 748–755. https://doi.org/10.1016/j.jesp.2011.02.011

Kogut, T. (2011b). The role of perspective taking and emotions in punishing identified and unidentified wrongdoers. *Cognition and Emotion*, *25*(8), 1491–1499. https://doi.org/10.1080/02699931.2010.547563

Kogut, T., & Ritov, I. (2005a). The "identified victim" effect: an identified group, or just a single individual? *Journal of Behavioral Decision Making*, *18*(3), 157–167. https://doi.org/10.1002/bdm.492

Kogut, T., & Ritov, I. (2005b). The singularity effect of identified victims in separate and joint evaluations. *Organizational Behavior and Human Decision Processes*, *97*(2), 106–116. https://doi.org/10.1016/j.obhdp.2005.02.003

Kogut, T., Ritov, I., Rubaltelli, E., & Liberman, N. (2018). How far is the suffering? The role of psychological distance and victims' identifiability in donation decisions. *Judgment and Decision Making*, *13*(5), 458–466. https://doi.org/10.1017/S1930297500008731

Krueger, F., & Hoffman, M. (2016). The emerging neuroscience of third-party punishment. *Trends in Neurosciences*, *39*(8), 499–501. https://doi.org/10.1016/j.tins.2016.06.004

Landmann, H., & Hess, U. (2017). What elicits third-party anger? The effects of moral violation and others' outcome on anger and compassion. *Cognition and Emotion*, *31*(6), 1097–1111. https://doi.org/10.1080/02699931.2016.1194258

Ledgerwood, A., Trope, Y., & Chaiken, S. (2010). Flexibility now, consistency later: Psychological distance and construal shape evaluative responding. *Journal of Personality and Social Psychology*, *99*(1), 32–51. https://doi.org/10.1037/a0019843

Lee, S., & Feeley, T. H. (2016). The identifiable victim effect: A meta-analytic review. *Social Influence*, *11*(3), 199–215. https://doi.org/10.1080/15534510.2016.1216891

Lee, S., & Feeley, T. H. (2018). The identifiable victim effect: Using an experimental-causal-chain design to test for mediation. *Current Psychology*, *37*(4), 875–885. https://doi.org/10.1007/s12144-017-9570-3

Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods*, *13*(3), 229–248. https://doi.org/10.18148/srm/2019.v13i3.7403

Leiner, D. J. (2024). *SoSci Survey* (Version 3.5.02) [Computer Software]. https://www.soscisurvey.de

Lewinsohn-Zamir, D., Ritov, I., & Kogut, T. (2017). Law and Identifiability. *Indiana Law Journal*, *92*(2), 505–555.

Loewenstein, G. F., Small, D., & Strnad, J. F. (2005). Statistical, identifiable and iconic victims and perpetrators. *SSRN*. https://doi.org/10.2139/ssrn.678281

Lotz, S., Baumert, A., Schlösser, T., Gresser, F., & Fetchenhauer, D. (2011). Individual differences in third-party interventions: How justice sensitivity shapes altruistic punishment. *Negotiation and Conflict Management Research*, *4*(4), 297–313. https://doi.org/10.1111/j.1750-4716.2011.00084.x

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.12558

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Mae Boag, E., & Wilson, D. (2013). Does engaging with serious offenders change students' attitude and empathy toward offenders? A thematic analysis. *The Journal of Forensic Psychiatry & Psychology*, *24*(6), 699–712. https://doi.org/10.1080/14789949.2013.841973

Maier, M., Wong, Y. C., & Feldman, G. (2024). Revisiting and rethinking the identifiable victim effect: Replication and extension of Small, Loewenstein, and Slovic (2007). *Collabra: Psychology*, *9*(1), 90203. https://doi.org/10.1525/collabra.90203

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, *52*(2), 464–488. https://doi.org/10.3758/s13428-019-01246-w

Majumder, R., Tai, Y. L., Ziano, I., & Feldman, G. (2022). *Revisiting the impact of singularity on the identified victim effect: An unsuccessful replication and extension of Kogut and Ritov (2005a) Study 2*. https://doi.org/10.17605/OSF.IO/9QCPJ

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Mentovich, A., Yudkin, D., Tyler, T., & Trope, Y. (2016). Justice without borders: The influence of psychological distance and construal level on moral exclusion. *Personality and Social Psychology Bulletin*, *42*(10), 1349–1363. https://doi.org/10.1177/0146167216659477

Mischkowski, D., Glöckner, A., & Lewisch, P. (2018). From spontaneous cooperation to spontaneous punishment – Distinguishing the underlying motives driving spontaneous behavior in first and second order public good games. *Organizational Behavior and Human Decision Processes*, *149*, 59–72. https://doi.org/10.1016/j.obhdp.2018.07.001

Moche, H., Gordon-Hecker, T., Kogut, T., & Västfjäll, D. (2022). Thinking, good and bad? Deliberative thinking and the singularity effect in charitable giving. *Judgment and Decision Making*, *17*(1), 14–30. https://doi.org/10.1017/S1930297500009001

Moche, H., Karlsson, H., & Västfjäll, D. (2024). Victim identifiability, number of victims, and unit asking in charitable giving. *PLoS ONE*, *19*(3), e0300863. https://doi.org/10.1371/journal.pone.0300863

Moche, H., & Västfjäll, D. (2021). Helping the child or the adult? Systematically testing the identifiable victim effect for child and adult victims. *Social Influence*, *16*(1), 78–92. https://doi.org/10.1080/15534510.2021.1995482

Moore, M. (2020). *Mechanical choices: The responsibility of the human machine*. Oxford University Press. https://doi.org/10.1093/oso/9780190863999.001.0001

Myers, B., & Greene, E. (2004). The prejudicial nature of victim impact statements: Implications for capital sentencing policy. *Psychology, Public Policy, and Law*, *10*(4), 492–515. https://doi.org/10.1037/1076-8971.10.4.492

Oceja, L. V., Heerdink, M. W., Stocks, E. L., Ambrona, T., López-Pérez, B., & Salgado, S. (2014). Empathy, awareness of others, and action: How feeling empathy for one-among-others motivates helping the others. *Basic and Applied Social Psychology*, *36*(2), 111–124. https://doi.org/10.1080/01973533.2013.856787

O'Mara, E. M., Jackson, L. E., Batson, C. D., & Gaertner, L. (2011). Will moral outrage stand up?: Distinguishing among emotional reactions to a moral violation. *European Journal of Social Psychology*, *41*(2), 173–179. https://doi.org/10.1002/ejsp.754

Oppenheimer, D. M., & Olivola, C. Y. (2011). *The science of giving: Experimental approaches to the study of charity*. Psychology Press. https://doi.org/10.4324/9780203865972

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Pałka, P., Bystranowski, P., Janik, B., & Próchnicki, M. (2023). When "a citizen" becomes little Mary J.: The abstract-concrete effects in legal reasoning and the rule of law. *Dritto & Questioni Pubbliche, Special publication*, 248–264.

Passarelli, T. O., & Buchanan, T. W. (2020). How do stress and social closeness impact prosocial behavior? *Experimental Psychology*, *67*(2), 123–131. https://doi.org/10.1027/1618-3169/a000482

Pedersen, E. J., McAuliffe, W. H. B., & McCullough, M. E. (2018). The unresponsive avenger: More evidence that disinterested third parties do not punish altruistically. *Journal of Experimental Psychology: General*, *147*(4), 514–544. https://doi.org/10.1037/xge0000410

Pfattheicher, S., Sassenrath, C., & Keller, J. (2019). Compassion magnifies third-party punishment. *Journal of Personality and Social Psychology*, *117*(1), 124–141. https://doi.org/10.1037/pspi0000165

Przepiorka, W., & Liebe, U. (2016). Generosity is a sign of trustworthiness—the punishment of selfishness is not. *Evolution and Human Behavior*, *37*(4), 255–262. https://doi.org/10.1016/j.evolhumbehav.2015.12.003

Qu, C., Tang, Z., Zhang, H., Hu, Y., & Dreher, J. C. (2018). Neural modulation of social distance on third-party punishment. *BioRxiv*, 274720. https://doi.org/10.1101/274720

Rim, S. Y., Hansen, J., & Trope, Y. (2013). What happens why? Psychological distance and focusing on causes versus consequences of events. *Journal of Personality and Social Psychology*, *104*(3), 457–472. https://doi.org/10.1037/a0031024

Roberts, J. V., & Hough, M. (2005). The state of the prisons: Exploring public knowledge and opinion. *The Howard Journal of Criminal Justice*, *44*(3), 286–306. https://doi.org/10.1111/j.1468-2311.2005.00373.x

Rudolph, U., Roesch, S. C., Greitemeyer, T., & Weiner, B. (2004). A meta-analytic review of help giving and aggression from an attributional perspective: Contributions to a general theory of motivation. *Cognition and Emotion*, *18*(6), 815–848. https://doi.org/10.1080/02699930341000248

Russell, P. S., & Giner-Sorolla, R. (2013). Bodily moral disgust: What it is, how it is different from anger, and why it is an unreasoned emotion. *Psychological Bulletin*, *139*(2), 328–351. https://doi.org/10.1037/a0029319

Sah, S., & Loewenstein, G. (2012). More affected = more neglected: Amplification of bias in advice to the unidentified and many. *Social Psychological and Personality Science*, *3*(3), 365–372. https://doi.org/10.1177/1948550611422958

Schein, C., & Gray, K. (2014). The prototype model of blame: Freeing moral cognition from linearity and little boxes. *Psychological Inquiry*, *25*(2), 236–240. https://doi.org/10.1080/1047840X.2014.901903

Schelling, T. C. (1968). The life you save may be your own. In S. B. Chase (Ed.), *Problems in public expenditure and analysis* (pp. 127–162). Brookings Institution.

Seip, E. C., Van Dijk, W. W., & Rotteveel, M. (2014). Anger motivates costly punishment of unfair behavior. *Motivation and Emotion*, *38*(4), 578–588. https://doi.org/10.1007/s11031-014-9395-4

Sifferd, K. L. (2024). Deserving blame, and sometimes punishment. *Criminal Law and Philosophy*, *18*(1), 133–150. https://doi.org/10.1007/s11572-023-09668-6

Small, D. A., & Loewenstein, G. (2005). The devil you know: The effects of identifiability on punishment. *Journal of Behavioral Decision Making*, *18*(5), 311–318. https://doi.org/10.1002/bdm.507

Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, *102*(2), 143–153. https://doi.org/10.1016/j.obhdp.2006.01.005

Squires, J. E. R. (1968). Blame. *The Philosophical Quarterly*, *18*(70), 54–60. https://doi.org/10.2307/2218028

Stallen, M., Rossi, F., Heijne, A., Smidts, A., De Dreu, C. K. W., & Sanfey, A. G. (2018). Neurobiological mechanisms of responding to injustice. *Journal of Neuroscience*, *38*(12), 2944–2954. https://doi.org/10.1523/jneurosci.1242-17.2018

Struthers, C. W., Eaton, J., Santelli, A. G., Uchiyama, M., & Shirvani, N. (2008). The effects of attributions of intent and apology on forgiveness: When saying sorry may not help the story. *Journal of Experimental Social Psychology*, *44*(4), 983–992. https://doi.org/10.1016/j.jesp.2008.02.006

Sudhir, K., Roy, S., & Cherian, M. (2016). Do sympathy biases induce charitable giving? The effects of advertising content. *Https://Doi.Org/10.1287/Mksc.2016.0989*, *35*(6), 849–869. https://doi.org/10.1287/MKSC.2016.0989

Twardawski, M., Gollwitzer, M., Pohl, S., & Bošnjak, M. (2022). What drives second- and third-party punishment? Conceptual replications of the “intuitive retributivism” hypothesis. *Zeitschrift Für Psychologie*, *230*(2), 77–83. https://doi.org/10.1027/2151-2604/a000454

Unnever, J. D., & Cullen, F. T. (2009). Empathetic identification and punitiveness: A middle-range theory of individual differences. *Theoretical Criminology*, *13*(3), 283–312. https://doi.org/10.1177/1362480609336495

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*(4), 301–325. https://doi.org/10.1023/a:1026231622102

Vu, L., Molho, C., Soraperra, I., Fiedler, S., & Shalvi, S. (2024). Giving (in) to help an identified person. *Journal of Experimental Social Psychology*, *110*, 104557. https://doi.org/10.1016/J.JESP.2023.104557

Wiss, J., Andersson, D., Slovic, P., Västfjäll, D., & Tinghög, G. (2015). The influence of identifiability and singularity in moral decision making. *Judgment and Decision Making*, *10*(5), 492–502. https://doi.org/10.1017/S1930297500005623

Yan, D., Sengupta, J., & Hong, J. (2016). Why does psychological distance influence construal level? The role of processing mode. *Journal of Consumer Research*, *43*(4), 598–613. https://doi.org/10.1093/jcr/ucw045

Yang, Q., Hoffman, M., & Krueger, F. (2024). The science of justice: The neuropsychology of social punishment. *Neuroscience & Biobehavioral Reviews*, *157*, 105525. https://doi.org/10.1016/j.neubiorev.2023.105525

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

Żuradzki, T. (2019). The normative significance of identifiability. *Ethics and Information Technology*, *21*(4), 295–305. https://doi.org/10.1007/S10676-018-9487-z