Law and Order:

The Timing of Mitigating Evidence Affects Punishment Decisions

Emily B. Conder [a], Christopher Brett Jaeger [b], & Jonathan D. Lane [a*]

[a] *Department of Psychology & Human Development, Vanderbilt University, Nashville, TN 37203*

[b] *Baylor Law School, 114 S. University Parks Dr., Waco, TX 76706*

*Corresponding Author:
Jonathan D. Lane
jonathan.lane@vanderbilt.edu

This is a pre-press draft of a manuscript published in *Thinking & Reasoning*. For the published version of this manuscript, please see:
https://doi.org/10.1080/13546783.2023.2172077

Abstract

When we hear about a transgression, we may consider whether the perpetrator's individual circumstances make their transgression more understandable or excusable. Mitigating circumstances may reduce the severity of punishment that is deemed appropriate, both intuitively and legally. But importantly, in courts of public opinion and of law, mitigating information is typically presented only after information about a perpetrator's transgression. We explore whether this sequence influences the force of mitigating evidence. Specifically, in two studies, we examined whether presenting evidence about a perpetrator's background before or after evidence of their violation influenced how severely U.S. participants punished perpetrators. In Study 1 (N=132), evidence about the perpetrator's mitigating circumstances reduced punishment only when it was presented before evidence about the perpetrator's violation. Study 2 (N=316) additionally revealed this moderating effect of presentation order across a variety of premeditated and impulsive violations. These findings are consistent with person-centered theories of punishment and with the Story Model of adjudication.

Keywords: transgressions; punishment decisions; extenuating circumstances; order effects

Law and Order:

The Timing of Mitigating Evidence Affects Punishment Decisions

Imagine that you are a school official deciding how to punish a student for stealing money and food from a classmate. As you weigh potential punishments, you learn that the student was raised in poverty and had endured abuse from their parents. Will knowledge of the student's circumstances affect how you punish them? Many studies have investigated variants of this question (e.g., Barnett et al., 2004; Robinson, Jackowitz, & Bartels, 2012; Bell et al., 2016; Meixner, 2022). The studies reported herein addresses a related, unexplored question: Would your punishment decision be different if you knew of the student's extenuating circumstances *before* you learned about their stealing?

The timing of evidence may play an important role in our everyday moral reasoning and decision-making. For instance, we judge friends and loved ones differently than acquaintances or strangers who commit the *same* violations (Forbes & Stellar, 2022). This is perhaps because, *prior* to learning about their violations, we possess more (generally positive) background knowledge about our friends and loved ones. Thus, what we know about others' prior experiences may soften the perceived magnitude of their violations or may partially excuse their violations. The current studies directly examine how the presentation of information about perpetrators' mitigating circumstances (before vs. after one learns about their violations) influences the punishment assigned to perpetrators. In what follows, we review work on associations between mitigating evidence and punishment decisions. We explain why the timing of evidence might matter, integrating a person-centered perspective on punishment (Landy & Uhlmann, 2018; Uhlmann, Pizarro, & Diermeier, 2015) with the Story Model of adjudication (Pennington & Hastie, 1991). This integrated perspective motivates the key hypothesis that we

test in the current studies, that the effect of mitigating circumstances on punishment decisions will depend on *when* evidence of mitigating circumstances is presented.

Punishment Decisions

Research on moral decision-making has identified many factors that can influence punishment decisions. Researchers have traditionally focused on factors that are temporally and causally related to the punishable act, particularly the *nature* of the act (e.g., firing a gun versus throwing a tennis ball), the *outcome* of the act (i.e., the harm caused), and the perpetrator's *mental states* (e.g., intent) during the act (Alicke, 1992; Cushman, 2008, 2015; Cushman et al., 2013; Young & Tsoi, 2013). These factors may influence punishment decisions in several ways. For example, punishment decisions might differ if a perpetrator's mental functioning is disrupted at the time of the offending act (e.g., due to an untreated brain tumor) or if they encounter circumstances that inspire uncontrollable passions (e.g., finding a partner with another lover) (Cushman, 2015). In these cases, participants tend to assign less punishment to perpetrators, who are viewed as not fully in control of their harmful behavior (e.g., Darley, Carlsmith, & Robinson, 2000; Robinson & Darley, 1995).

Other work demonstrates that punishment decisions may be informed by broad assessments of perpetrators' moral qualities. This "person-centered" account of moral decision-making stresses that humans are motivated to quickly (and often, spontaneously) draw broad inferences about a perpetrator's moral characteristics (Landy & Uhlmann, 2018; Uhlmann et al., 2015; Winter & Uleman, 1984; Hamlin, Wynn, & Bloom, 2007). These broad evaluations can then affect decision-makers' inferences about a perpetrator's causal responsibility for harms (Alicke, 1992) and their intent (Nadler, 2012; Nadler & McDonnell, 2011), as well as decision-

makers' evaluation of how much punishment a perpetrator deserves (for review, see Uhlmann et al., 2015).

Person-centered notions of "just desert" are central to judgments about justice and punishment (e.g., Darley, Carlsmith, & Robinson, 2000; Darley 2009; Heuer, 2005; Heuer, Blumenthal, Douglas, & Weinblatt, 1999). Drolet, Hafer, and Heuer (2016) found that participants were more tolerant of an interrogation target being tortured if the target's past behavior (unrelated to the current interrogation) was more reprehensible (e.g., killing children) versus less reprehensible (e.g., injuring soldiers) (see also Carlsmith & Sood, 2009). The association between the target's past behavior and participants' tolerance of the target's torture was mediated by participants' belief that the target deserved harsh treatment. With respect to punishment, Nadler (2012) found that participants assigned greater punishment to a reckless skier who injured another person when the skier was described as having poor character (e.g., often late to work, loafs around town, rarely helps his family) versus when the skier was described as having good character (e.g., a model employee, volunteers at an animal shelter, regularly helps his family). Conversely, evidence that a perpetrator endured a difficult past may soften the punishment assigned to them for their offenses (e.g., Barnett et al., 2004; Robinson et al., 2012; Bell Holleran et al., 2016; Meixner, 2022; but see Appelbaum & Scurich, 2014; Berryessa, 2016; Lynch et al., 2019). In sum, what decision-makers know about a perpetrator's background can sway their punishment decisions.

The Role of Timing

In the current studies, we explore the possibility that the effect of mitigating evidence on punishment decisions varies based on the *order* in which decision-makers are presented evidence

about the perpetrator's violation and the perpetrator's mitigating circumstances. There are at least two theoretical motivations to expect this order effect.

The first motivation is the person-centered perspective on punishment. As noted above, a person-centered perspective implies that evidence about a perpetrator's background circumstances influence how others punish them. It also implies that the *timing* of such background evidence should matter, because evidence of a perpetrator's illicit acts function as a source of information about their global moral character. The extent to which the information is interpreted as diagnostic of moral character depends on context (for review, see Uhlmann et al., 2015). When the only information available about a perpetrator concerns their violative act, we may be prone to draw initially negative global inferences about their character. Thus, the timing of mitigating evidence should matter: learning about a perpetrator's background circumstances *before* learning about their illicit act should add nuance to inferences about the perpetrator's character *before* negative global inferences can be drawn from their violation alone.

A second motivation for the prediction that the effects of mitigating evidence will vary based on its timing follows from theories of adjudication prominent in the literature on legal decision making. We use the term adjudication to mean "the formal process by which litigants offer evidence to a decisionmaker (typically a juror) charged with finding facts and applying rules that specify the significance of such facts" (Kahan, 2015, p. 56). Much research at the intersection of psychology and law has used adjudication as a context to explore how decisionmakers use facts to draw inferences and make evidence-based decisions. A general theme that has emerged from this research is that decisionmakers do not process evidence like classical Bayesians; that is, they do not weigh pieces of evidence independently and update their assessment of a case in an algebraic manner. Rather, they tend to *integrate* evidence with

existing knowledge to construct and test narratives about what happened in a given case (Bilz, 2010; Feigenson, 1995; Holyoak & Simon, 1999; Kahan, 2015; Pennington & Hastie, 1986, 1990, 1992).

Perhaps the most-studied narrative framework is Pennington and Hastie's "Story Model." The Story Model posits that legal decisionmakers use the evidence presented during adjudication, together with their existing knowledge about the world and about story structure, to construct and compare possible narratives of what happened in a case. Decisionmakers come to cases equipped with *stock stories* (e.g., 'bullied student snaps'; 'wife poisons unfaithful husband') that function as templates for organizing, interpreting, and filling gaps in the evidence presented to them (Pennington & Hastie, 1986). This narrative-construction process is ongoing throughout adjudication, as decisionmakers attempt to integrate evidence into coherent stories (Feigenson, 1995; Kahan, 2015). Ultimately, decisionmakers choose what they consider to be the most complete, coherent story and render their legal decision based on that story. The Story Model predicts that the order in which evidence is presented will likely affect legal decisions. In particular, evidence that decisionmakers encounter *earlier* will have *outsized* influence as it shapes the decisionmakers' *initial* selection of story templates (Kahan, 2015; Pennington & Hastie, 1988, 1991, 1992). Whether a decisionmaker learns about the defendant's crime vs. the defendant's difficult life circumstances *first* might cause them to select very different story templates, leading to different conclusions.

Thus, the person-centered perspective of moral decision-making and the Story Model of adjudication both inspire the prediction that the effect of mitigating evidence on decision-makers' subsequent punishment judgments will depend on its timing (relative to when evidence of an actor's violation is presented). The current studies test this general hypothesis. To our

knowledge, no studies that have examined punishment decisions have manipulated *when* evidence of mitigating circumstances is introduced.

Current Studies

In two studies, we tested whether providing information about perpetrators' background circumstances *before or after* information about the perpetrators' violations would moderate the effect of background circumstances on punishment decisions. Study 1 was an initial exploration of this question: participants were asked to assign punishment to perpetrators who had either *negative* or *neutral* backgrounds, and perpetrators' background information was presented either before or after information about their violation. Participants read a series of vignettes, each consisting of two components: a description of the violation (e.g., stealing or injuring someone) and a description of the perpetrator's background circumstances. In some vignettes, the perpetrator's background circumstances were *negative* (e.g., growing up in poverty, or being socially excluded). In other vignettes, the perpetrator's background circumstances were comparatively *neutral* (e.g., growing up middle-class, or making new friends). We use the term "negative" not to reflect our own evaluations of these conditions, but as shorthand to refer to positions that are comparatively less privileged, less supportive, or that people often wish to avoid. Key words included in the "negative" scenarios (e.g., "mean"; "poverty"; "injury") are typically interpreted as negatively-valanced (see Warriner, Kuperman, & Brysbaert, 2013). We use the term "neutral" to refer to expected outcomes (e.g., completing a project as anticipated), normative conditions for the perpetrator (e.g., growing up middle-class) or conditions only tangentially related to the perpetrator (e.g., reading coworker's derogatory comments about celebrities). For each study, we included a variety of background circumstances and a variety of transgressions to evaluate the robustness of these effects.

We hypothesized that participants evaluating perpetrators with mitigating/negative circumstances would assign less severe punishment than participants evaluating perpetrators with *neutral* circumstances; however, *only* when the mitigating/negative circumstances were presented *before* information about the perpetrators' transgressions. Inspired by previous work, for exploratory purposes, we also evaluated other factors that might influence punishment decisions directly or by mediating the effects of background, including participants' inferences about perpetrators' character and agency, and participants' sympathy and empathy toward the perpetrators (see Alicke, 1992; Nadler, 2012; Uhlmann, et al., 2015). Study 1 revealed the anticipated order effect: the order in which participants learned of perpetrators' background circumstances (mitigating or neutral) moderated the relation between perpetrators' background circumstances and participants' punishment decisions. Study 2 was designed to test the replicability of this effect and to test its boundary conditions, evaluating whether it manifests for impulsive transgressions, premeditated transgressions, or both. As in Study 1, perpetrators' violations were described either before or after descriptions of their backgrounds. Study 2 included an additional (positive) background condition, to help further specify the influence of perpetrators' background circumstances on punishment decisions.

All study methods, materials, and consent documents were approved by Vanderbilt University's Institutional Review Board ("Accounting for Extenuating Circumstances in Moral Judgements", IRB# 170239).

Study 1

Method

Participants

Participants (N = 132) were recruited using Amazon's Mechanical Turk. An additional

five participants were excluded for failing to correctly answer at least five of the six reading-comprehension questions about the violation scenarios (probability of randomly answering 5-6 of the 6 questions correctly, $p = .005$). The target sample size was informed by an a priori power analysis using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007), which assessed the number of participants needed to detect medium effect sizes (e.g., $fs \geq .25$) with statistical power at the recommended .80 level and $\alpha = .05$ (Cohen, 1992). This analysis identified a minimum recruitment goal of 116 participants.

The survey was open online to all Mechanical Turk (MTurk) workers in the United States who were at least 18 years in age and spoke English. Data collected through Mturk have generally been found to be at least as reliable as data obtained through traditional recruitment methods, and to represent a significantly more diverse group of participants than conventional samples of college students (Behrend et al., 2011; Buhrmester, Kwang, & Gosling, 2011; Kees, Berry, Burton, & Sheehan, 2017). TurkPrime.com was used to administer the survey, and several methods were employed to ensure that each participant was unique (e.g., allowing only one participant per IP address; to maintain anonymity, participants' IP addresses were not stored in data files). To ensure that participants were legitimate, we required that at least 95% of each participant's prior Mturk 'work' had been approved, employed comprehension-check questions, CAPTCHA, and a survey completion confirmation code. Participants were compensated $1 for completing the survey on Qualtrics.com (average duration was 13 minutes, including the consenting phase). After completing the informed consent, participants were asked to report their age (M_{age} = 31.67 years, SD_{age} = 9.40 years, range = 19-66 years), gender identification (77 male, 54 female, 1 "preferred not to say"), and zip code (participants could opt to skip any of these questions).

After completing the focal portion of the survey (i.e., after reading and answering questions about six transgression scenarios, described later), all participants completed an additional voluntary demographic questionnaire, which we used to further characterize our sample. Most participants (98%) reported that their first language was English. The majority (65%) identified as exclusively White/Caucasian, 14% identified as exclusively Black/African-American, 10% exclusively Asian/Asian-American, 7% identified as Hispanic/Latino White, 4 participants identified with multiple ethnicities (Asian/White, Black/White, Other/White), and 3 participants identified as Other (one of whom also identified as Hispanic/Latino). Three participants did not report whether they identified as Hispanic/Latino. In terms of education, 40% of participants reported having earned a bachelor's degree, 38% had completed some college, 12% earned a high school diploma, 7% earned a master's degree, 2% earned a doctorate, and 1% reported completing some high school. Additional demographic data (religious identity, religious participation, and political identity) are presented in Supplementary Materials.

Stimuli

Each participant read six vignettes about different moral violations. Each vignette consisted of two passages, one with information about the violation and one with information about the perpetrator's background circumstances. Our goal in creating these vignettes was to eventually study the influence of extenuating circumstances on both adults' and children's judgments of perpetrators. Thus, all backgrounds and violations were designed to be appropriate for research with participants under the age of eighteen, and we did not include some of the more extreme background circumstances (e.g., child abuse) or violations (e.g., murder) that have been used in past work with adults (for examples, see Appelbaum & Scurich, 2014; Barnett, Brodsky & Davis, 2004; Robbins & Litton, 2018). The violation and background passages were written so

that they could be mixed and matched randomly—i.e., any violation passage could be paired with any background passage, in any order, to create a coherent vignette. Randomization features from Qualtrics.com were used to randomize background-violation pairings for each participant.

Violations. The violation passage of each vignette described a character committing one violation that resulted in a clear, negative outcome. The six violations included: disturbing the peace, destruction of property, selling secrets, theft, online bullying, and battery (the perpetrator tripped a coffee shop patron with their leg). Each violation was loosely based on real events described in online news articles. For the full text of each violation, visit: https://osf.io/gf5rb/?view_only=5439416be8dc4ec88e3354409882d1c3.

We created two versions of each of the six violation passages: a more *impulsive* version and a more *premediated* version. The two versions differed in the extent to which, and duration over which, the perpetrator planned the violation. For example, in one passage, the perpetrator uses their leg to trip and injure a stranger who repeatedly cuts in line at the perpetrator's local café. In the premeditated version, the perpetrator decided they "would do something about it the next time it happened" days before ultimately tripping the stranger with their leg. Whereas, in the impulsive version, the perpetrator "feels a surge of anger and suddenly sticks out their leg and trips the stranger." Using these violations in a pilot study, we found that participants did indeed attribute more *intent* to perpetrators who committed premeditated violations than to those who committed impulsive violations (see Supplementary Materials for a brief description of this pilot study and average intent scores by violation type). Violation type (premeditated vs. impulsive) was manipulated within-subjects, such that each participant read about three *premeditated* violations and three different *impulsive* violations.

Backgrounds. The personal *background* of each perpetrator was manipulated between-subjects (negative circumstance or neutral circumstance), and participants were randomly assigned using Qualtrics.com's randomization features. Approximately half of the participants (*n* = 64) read about perpetrators with *negative* background circumstances and the other half (*n* = 68) read about perpetrators with *neutral* background circumstances. The six *negative* backgrounds described the perpetrator as either growing up in poverty, being cyberbullied, being socially excluded, exerting effort with no payoff, having a genetic disorder, or having a traumatic brain injury. The six *neutral* backgrounds matched the negative backgrounds in length and general topic and either described the perpetrator as growing up middle class, visiting websites about celebrities, making friends in a new city, having effort pay off as expected, applying for health insurance, or receiving a blood test. The full versions of all 12 backgrounds can be found here: https://osf.io/gf5rb/?view_only=5439416be8dc4ec88e3354409882d1c3.

Procedure

Participants read a general description of how the survey would proceed—they would read six vignettes and answer five questions following each vignette. Participants were urged to read each vignette carefully, to "answer each question based on how you would react to the situation if it were to happen in real life," and to, "consider each scenario independently." The perpetrator in each vignette had a different name and the names were gender-matched to each participant. The one participant who did not indicate their gender read about characters with male names. The procedure for each vignette was as follows:

Each vignette was initially presented in two parts, the violation passage and the background passage, displayed separately on successive web pages. The presentation order (violation-first or background-first) was randomly assigned between subjects. Participants

evaluated perpetrators' *character*, assigned *punishment* to perpetrators, reported their *sympathy* and *empathy* toward perpetrators, and provided their estimates of perpetrators' *control*. Participants evaluated the perpetrator's *character* twice—once after the violation passage and once after the background passage—using a 7-point Likert scale (from 1 = "very bad" to 7 = "very good"). Next, the entire vignette (i.e., both the violation passage and the background passage) appeared on the same page (arranged in the same order as previously presented) along with a prompt asking participants to rate how severely that perpetrator should be *punished* for the specific violation (from 1 "not at all punished" to 7 "severely punished"). For exploratory purposes, participants were asked how much *control* they believed the perpetrator had over the situation (from 1 = "no control" to 7 = "complete control"), how much the participant could "relate to" the perpetrator (from 1 = "not at all" to 7 = "very much"; a measure of *empathy*), and how "sorry" they felt for the perpetrator (from 1 = "not at all" to 7 = "very much"; a measure of *sympathy*). After answering all of these questions, participants were shown a new web page with a multiple-choice question to assess whether the participant read and comprehended the vignette (see Supplementary Materials for example multiple-choice comprehension questions). This procedure was repeated for each of the six vignettes. After reading and responding to all vignettes, participants were asked to voluntarily provide demographic information.

Results

Punishment Decisions

The primary purpose of Study 1 was to investigate the relation between perpetrators' backgrounds (negative circumstances vs. neutral circumstances) and the severity of punishment assigned to perpetrators, based on the order in which background information was presented. We examined participants' punishment decisions using a mixed-effects ANOVA that included

Background type (negative circumstances vs. neutral circumstances) and Presentation order (background before violation vs. after violation) as between-subjects factors, and Violation type (premeditated vs. impulsive) as a within-subjects factor. This analysis revealed a significant interaction of Background type and Presentation order ($F(1,128) = 7.72$, $p = .006$, $\eta_p^2 = .06$), depicted in Figure 1. Post hoc tests (Bonferroni adjusted $\alpha = .025$) revealed that participants who were presented with background information *before* violation information judged perpetrators as deserving less punishment if the perpetrators had experienced *negative* circumstances than if they had experienced *neutral* circumstances ($p = .016$). However, participants who were presented with background information *after* violation information assigned similar punishment to perpetrators who experienced *negative* circumstances and perpetrators who experienced *neutral* circumstances ($p = .138$). There were no other significant effects (see Table S1 in Supplementary Materials for the complete ANOVA results). Analyses involving participants' evaluations of perpetrators' character, sympathy and empathy toward perpetrators, and inferences of perpetrators' control are included in Supplementary Materials.

Discussion

In Study 1, we found evidence of the predicted order effect. Participants presented with information regarding perpetrators' negative circumstances punished perpetrators less severely *only* if information about perpetrators' circumstances was presented *before* information about their violations. When information about perpetrators' circumstances was presented *after* information about their violations, it did *not* influence participants' punishment decisions. In other words, participants' punishment decisions *were* mitigated when they had information about perpetrators' extenuating circumstances *before* learning about their violations. Punishment

decisions were *not* mitigated if the information about perpetrators' extenuating circumstances were introduced later.

We interpret Study 1 findings as evidence that information about negative circumstances *reduces* how severely decision-makers choose to punish perpetrators, *if* such evidence is presented first. This finding inspires an additional question—are *positive* circumstances an exacerbating force, that would yield the opposite order effect? Perhaps greater punishment would be administered to perpetrators whose positive circumstances are described before vs. after their violations. To explore this possibility, we created an additional *positive* version of each background passage for Study 2. For example, for background passages about a perpetrator's economic circumstances, participants could be presented with a passage that described *negative* circumstances (e.g., grew up in poverty), *neutral* circumstances (e.g., grew up middle class), or nominally *positive* circumstances (e.g., grew up wealthy). As noted earlier, labels such as "positive" and "negative" do not reflect our own evaluations of these circumstances but reflect normative evaluations of the circumstances—e.g., words used in the "negative" scenarios (e.g., "mean"; "poverty"; "injury") tend to be interpreted as negatively-valanced; whereas key words used in the "positive" scenarios (e.g., "like"; "wealthy"; "excellent") tend to be interpreted as positively-valanced (see Warriner et al., 2013).

The Role of Violation Type

In Study 1 we included both premeditated and impulsive moral violations (manipulated within subjects). This allowed us to test whether order effects existed across a wide set of scenarios. In Study 2, we were interested in evaluating the replicability of the order effect identified in Study 1 and testing the boundary conditions for that effect. To eliminate the possibility of carry-over effects between the conditions, separate samples of participants were

recruited for Study 2—one group made punishment decisions about *premeditated* violations and the other group made decisions about *impulsive* violations.

Initially, two separate studies were conceptualized and conducted, differing only in whether participants reasoned about impulsive violations (pre-registered: http://aspredicted.org/blind.php?x=8md3uk) or premeditated violations (pre-registered: http://aspredicted.org/blind.php?x=rw2zt7). Each study had a target sample of 158 participants, informed by *a priori* power analyses using G*Power 3.1 (Faul et al., 2007) to detect medium-sized effects (e.g., fs $\geq .25$) with power $\geq .80$ and $\alpha = .05$. Participants in one study could not participate in the other. Results for the individual pre-registered studies are presented in Supplementary Materials. For the sake of parsimony and to increase our statistical power, we combined the two datasets so that we may additionally test for main effects and interaction effects involving the type of violation (premeditated vs. impulsive, treated as a between-subjects variable). As the two samples were recruited at different points in time, conceivably they might differ demographically. However, the two samples were equivalent in age ($t(314) = -.82$, $p = .42$), education ($t(313)=-.65$, $p = .51$), distribution of males and females (Mann-Whitney U = 11358.5, $Z = -1.08$, $p = .28$), and proportion of participants who identified as Black/African-American (Mann-Whitney U = 12245, $Z = -.62$, $p = .53$), White/Caucasian (Mann-Whitney U = 12166, $Z = -.62$, $p = .54$), Asian/Asian-American (Mann-Whitney U = 12166, $Z = -.85$, $p = .40$), Native American (Mann-Whitney U = 12245, $Z = -1.15$, $p = .25$) and Hispanic/Latino (Mann-Whitney U = 12173, $Z = -.63$, $p = .53$).

Study 2

Method

Participants

Participants (N = 316), who had not participated in Study 1, were recruited using Mechanical Turk (MTurk). An additional 29 participants were excluded from this sample for failure to correctly answer all 4 of the multiple-choice reading comprehension questions (probability of randomly answering all 4 questions correctly, p = .004). Three participants were excluded for stopping and then re-starting the survey.

The survey was open to all MTurk workers in the United States who were at least 18 years old, spoke English, and did not participate in Study 1. TurkPrime.com was used to administer the survey, and we employed the same methods as in Study 1 to ensure that each participant was unique and submitted legitimate data. Each participant was compensated $1 for completing the survey on Qualtrics.com, which took an average of 9 minutes. After providing informed consent, participants were asked to report their age (M_{age} = 36.61, SD_{age} = 11.58 years, range = 19-74 years), gender (58% male, 40% female, 1% preferred not to say, and 1 participant responded “other”), and zip code.

All participants completed a voluntary demographic questionnaire. Most participants (98%) reported that their first language was English. Most participants (76%) identified as exclusively White/Caucasian, 7% identified as exclusively Asian/Asian-American, 6% identified as exclusively Black/African-American, 1 participant identified as exclusively Native-American, 2 as exclusively Hispanic/Latino, and an additional 9% of participants identified with multiple ethnic or racial categories (Native-American/White, Black/White, and Hispanic/Latino Native-American/Other/White). Three participants did not report whether they identified as

Hispanic/Latino. With respect to education, 41% of participants reported earning a bachelor's degree, 34% completed some college, 17% had earned a high school diploma, 7% had earned a master's degree, and 1% reported earning a doctorate (one participant did not report their education). Additional demographic data (religious identity, religious participation, and political identity) are presented in Supplementary Materials.

Stimuli

For Study 2, we reduced the total number of vignettes from six to four. We removed two *violation* passages that were conceptually similar to other violation passages (one of two that had involved destruction of property and one of two that had involved theft). We also removed two *backgrounds*—one background of a pair that were conceptually similar (both involved peer victimization) and for which participants in Study 1 had assigned similar punishment ratings, and one background that involved a perpetrator's genetic disorder, based on past research suggesting that genetic evidence might not influence punishment decisions as strongly as other mitigating circumstances (Appelbaum & Scurich, 2014; Berryessa, 2016).

Violations. One goal of this study was to evaluate whether the key interaction of Presentation order and Background type differed for *impulsive* and *premeditated* violations. Using the remaining four violation passages from Study 1, minor edits were made to enhance the spontaneity of each offense for *impulsive* violations, and to further highlight the perpetrator's planning for *premeditated* violations. Additional minor edits were made to the passages to increase their relevance to the participants by including language that implied the violations occurred *recentl*y (e.g., changing "One day…" to "Last week…") and by changing the concluding phrases in each violation to present tense (rather than past tense). The Study 2 violation passages can be found here:

https://osf.io/gf5rb/?view_only=5439416be8dc4ec88e3354409882d1c3. Half of the participants (*n* = 158) were assigned to read about *impulsive* violations and the other half (*n* = 158) were assigned to read about *premeditated* violations.

Backgrounds. In Study 2 we addressed the possibility that *positive* circumstances might lead participants to punish perpetrators more severely for transgressing in the absence of difficult life experiences. To test this possibility, we created four *positive* background passages. Participants were randomly assigned to read about perpetrators who had experienced either negative (*n* = 106), neutral (*n* = 106), or positive (*n* = 104) circumstances. All three versions of each background passage focused on similar topics and were of similar length (for these passages, see https://osf.io/gf5rb/?view_only=5439416be8dc4ec88e3354409882d1c3).

Procedure

The procedure was similar to Study 1, but involved several revisions. First, instructions were revised slightly to increase both their ecological validity and relevance to participants. Specifically, participants were instructed at the beginning of the survey that "these scenarios are based on real events and real people" and that they should "respond to each question as if you are making decisions that will affect these people in real life."

For each of the four vignettes, participants read the first passage (either the violation passage or background passage) on one web page and read the second passage on a subsequent web page. After reading both passages separately, participants were shown both passages on the same screen and asked how severely they would punish the perpetrator for committing the violation (from 1 "not at all punished" to 7 "severely punished"). After reporting their punishment decision, participants answered a multiple-choice question to assess whether they read and comprehended each passage. Given that the primary purpose of Study 2 was to examine

punishment decisions, and given that in Study 1 our additional variables (e.g., character, empathy) did *not* moderate relations between perpetrators' background circumstances and punishment decisions, we did not include these additional measures. Presentation order (violation first or background first), and Background type (negative, neutral, or positive) were manipulated between-subjects.

Results

We evaluated how our focal variables influenced punishment decisions with an ANOVA that included Background type (negative vs. neutral vs. positive), Presentation order (background before vs. after violation), and Violation type (impulsive vs. premeditated). See Table S9 for full ANOVA results. This analysis revealed the hypothesized interaction of Presentation order and Background type ($F(2, 304) = 3.70$, $p = .026$, $\eta_p^2 = .024$), depicted in Figure 2. According to post-hoc pairwise comparisons, when participants were presented with the perpetrator's background *before* their violation (but not when they were presented with the violation first), there were significant differences in punishment severity across background types. When presented with the perpetrator's background before their violation, participants assigned the perpetrator *less severe* punishment if they had experienced *negative* backgrounds vs. *positive* backgrounds ($p < .001$) or *neutral* backgrounds ($p = .015$).

An additional way to explore the interaction of Presentation order and Background type is by comparing punishment judgments for the different Presentation orders (violation first vs. background first) when the perpetrators' backgrounds were *negative*, *neutral*, or *positive*. This post-hoc analysis revealed an effect of Presentation order only for *negative* backgrounds—participants administered less severe punishment when the perpetrator's background information was presented *before* their violation information ($p = .01$). No order effects were found when the

perpetrator had a *neutral* (p = .633) or *positive* background (p = .339). Thus, relative to all other conditions, participants administered the least severe punishment when they read about perpetrators' negative background circumstances before reading about their violations. Importantly, this effect was robust—as depicted in Tables S7 and S8, the effect was evident across all four violations and for three of the four backgrounds.

Although less pertinent to our study aims, the same ANOVA also revealed a significant effect of Violation Type ($F(1, 304) = 5.11$, p = .024, η_p^2 = .017), which was moderated by Background Order ($F(1, 304) = 4.33$, p = .038, η_p^2 = .014). Participants assigned greater punishment to perpetrators who committed the more premeditated (vs. impulsive) violations. However, this was only the case when violation information was presented before background information (premeditated: M = 4.91, SE = .12; impulsive: M = 4.38, SE = .12; $p < .01$), and not when violation information was presented after background information (premeditated: M = 4.57, SE = .12; impulsive: M = 4.55, SE = .12; p = .899).

General Discussion

The current studies are the first, to our knowledge, to examine whether the time at which information about perpetrators' background circumstances is presented influences how severely decisionmakers punish those perpetrators. In Study 1, participants who read about perpetrators' mitigating background circumstances (e.g., having been raised in poverty), punished perpetrators *less severely* than participants who read about perpetrators' neutral background circumstances (e.g., having been raised middle class) only if information about perpetrators' mitigating circumstances was presented *before* information about their violations. Information about mitigating circumstances did *not* influence punishment decisions if it was presented *after* violation information. In Study 2, we evaluated the replicability of the order effect detected in

Study 1 and tested the boundary conditions of the effect. When presented with background information first, participants' punishment decisions were less harsh for perpetrators with negative backgrounds compared to perpetrators with neutral or positive backgrounds; this was true for impulsive and premeditated violations. As in Study 1, when the perpetrators' background information was presented after information about their violations, the background information had *no effect* on participants' punishment decisions. This interaction effect seems quite robust—it was identified in two studies, and across various types of premeditated and impulsive violations (see Table S7). Although the two samples of participants in Study 2 were not *randomly* assigned to reason about 'premeditated' vs. 'impulsive' violations (a gold standard of experimental research that helps to ensure equivalent samples), post-hoc analyses revealed that the two samples were quite similar demographically.

Our findings are consistent with the person-centered theory of moral decision-making and the Story Model of adjudication. We observed that information about a perpetrator's mitigating circumstances predicted participants' punishment judgments. However, this was true only when mitigating evidence was introduced *before* information about the perpetrator's violation. These findings suggest that an initial intuition that someone deserves a particular punishment, given information about their violation, shapes whether and how much subsequent mitigating evidence about the person's circumstances is weighed. Specifically, one's early-formed intuitions that a perpetrator deserves significant punishment for a violation may reduce one's consideration of later-provided mitigating evidence. Likewise, the Story Model can account for the fact that, in Study 2, violation premeditation influenced punishment decisions only when *violation information* was presented *first*, before information about perpetrators' backgrounds.

Future research might further probe the link between the order effects we have observed and the theories that inspired our predictions. For example, the Story Model suggests that participants organize and remember the evidence presented to them in narrative form and thus information presented earlier has outsized influence. Future research could further test this account by explicitly probing participants' construction of story models and the relations among those models, presentation order, and punishment decisions. For example, after reviewing both background and violation information, participants might respond to open-ended prompts to summarize the evidence as they understand it. This would allow researchers to observe the relative placement and prominence of mitigating circumstances in *participants'* mental narratives of events, and to observe associations with presentation order and punishment decisions.

One theoretical divide that could be probed relates to the distinction between person-centered and act-centered accounts of punishment. Person-centered theories of punishment are grounded in virtue ethics, whereas act-centered theories (such as deontology) focus on moral evaluation of discrete acts without reference to the particular people committing them (Landy & Uhlman, 2018; Uhlmann et al., 2015). Our findings, along with much prior research, support the notion that person-centered factors may powerfully influence punishment judgments (Nadler, 2012; Uhlman et al., 2015). But the *relative* weight of person-centered factors and act-centered factors in punishments could be usefully explored in work that systematically varies the severity of the offense and the salience of mitigating—or exacerbating—circumstances. For example, researchers could present participants with transgressions of varying severity (e.g., murder, armed robbery, shoplifting, trespassing, jaywalking) counterbalanced with background information about the perpetrator's past transgressions of varying severity (e.g., they had previously committed the act of murder, armed robbery, shoplifting, trespassing, or jaywalking,

without prosecution). The relative effect sizes of current crime severity and background information would provide insight into the relative weight afforded act-centered and person-centered information.

One important caveat to note is that the order effects identified in the current studies are found for participants' reasoning about written vignettes; work is needed to test whether these effects generalize to real-world contexts and real court cases. Future work may evaluate how jurors' exposure to public discourse and mass media about particular cases and defendants influences their construction of mental narratives of events and jurors' subsequent punishment judgments. In future work focused on potential legal applications, more ecologically valid stimuli could be devised, including mock transcripts, videos with evidence summaries (e.g., Chao & Santos, 2019), or even full mock trials, with violations more typical of those prosecuted in the criminal legal system. Because the violations presented in the current studies were, by design, relatively mild, one may predict that punishment of more severe violations (e.g., homicide) would be unaffected by presenting mitigating evidence before violation information. Notably, in the current studies perpetrators' negative background circumstances were *also* relatively mild and did not include the more horrific circumstances that some defendants experience (e.g., child abuse). We anticipate that for more severe violations (e.g., homicide) similar order effects might exist when paired with evidence of severe mitigating circumstances (e.g., the perpetrator's history of being abused).

While we stress that more research is needed to bridge our survey findings to the real-world criminal legal system, assume for the moment that future research does ultimately build a sturdy bridge: What might be the implications of similar findings in more ecologically valid contexts? Many have argued that laws should reflect commonplace moral intuitions, and often

they do (e.g., Bilz & Nadler, 2009; Kelman, 2013). The fact that the law explicitly contemplates that courts will consider evidence about a defendant's individual circumstances in criminal sentencing reflects a policy decision that such evidence *should* matter (18 U.S.C. § 3553). Yet the way the legal system actually incorporates such evidence may undercut its effectiveness. When a case goes to trial in the American criminal legal system, typically a jury decides whether the defendant is guilty and (if so) the judge subsequently sets the sentence. Generally, mitigating evidence is introduced only at the sentencing phase, after guilt has been determined. At the sentencing phase, the judge is free to consider broad factors in determining the offender's punishment, including "the nature and circumstances of the offense" and the "history and characteristics of the defendant" (18 U.S.C. § 3553). Thus, lawyers at this phase are far freer to introduce evidence about the defendant's personal characteristics and background than they are during trial (e.g., Meixner, 2022). The consequence is that the judge typically hears the details of the defendant's particular (alleged) violation before learning about the defendant's mitigating individual circumstances during the sentencing phase. An implication of our findings (if reinforced in future work) is that, by the sentencing stage, evidence about the defendant's extenuating circumstances may do little to influence punishment decisions, even though such evidence is allowed because *it should* influence punishment decisions.

It is conceivable that judges, with the benefit of years of training and professional experience, are less prone to the observed order effect than are lay participants. Perhaps judges, can mentally compartmentalize evidence that is relevant only to the defendant's guilt from evidence that is relevant to the court's sentencing decision. However, research conducted with judges indicates that, while their training and experience help to reduce the *magnitude* of some cognitive biases, they remain very much susceptible to typical human cognitive limitations and

tendencies (Guthrie et al., 2000; Rachlinski & Wistrich, 2017). Thus, we would be surprised to observe order effects in mitigating evidence for lay decision-makers but not for judges, at least with respect to vignette-based studies. One could test these hypotheses in future research with samples of judges, although such samples are typically difficult to recruit.

Now, assume that the order effect we have described does manifest in the criminal legal system. If policymakers desire that evidence of mitigating circumstances play a greater role in sentencing, arguably affording more individualized justice, then they could devise new processes to allow the decisionmaker tasked with sentencing a defendant to learn about the defendant's mitigating circumstances *before* learning about the defendant's crime. For example, in a case where a jury decides the defendant's guilt and the judge decides the sentence, it is possible that the judge could hear or review limited information about the defendant's personal circumstances—at least the subset of personal circumstances not immediately related to the criminal act—earlier in the proceedings, before hearing about the defendant's alleged offense in great detail. A caveat is that this might create new risks of bias, with evidence of mitigating circumstances unduly affecting judicial decisions on pre-trial motions, the admissibility of evidence, and so on. But policymakers can balance these risks with the risks of systematically underweighting mitigating evidence that may be inherent to the current approach. One could also imagine a system where criminal cases are more rigidly bifurcated across decision-makers, with the sentence set by a decision-maker or decision-makers who review the relevant information but are not involved in the initial guilt determination. Indeed, at least one federal judge dreamed of a system where "the court's only function would be the determination of innocence or guilt and a board composed of persons trained and experienced in the problems of criminology would determine the nature and duration of corrective treatment" (Levin, 1967, p. 499).

Thus far, we have discussed potential implications for cases in which the criminal defendant elects to go to trial. However, in modern criminal practice, the majority of criminal defendants accept a plea deal rather than go to trial. After pleading guilty, the defendants are then sentenced by the judge, who may not know much about the defendant's offense before the sentencing stage (Meixner, 2022). This suggests an interesting natural experiment. If the weight judges afford mitigating evidence in their punishment decisions depends on what they have previously learned about the defendant's offense, one would expect mitigating evidence about a defendant's personal circumstances to factor more heavily in judges' sentencing decisions in cases involving plea deals (where judges hear less about the details of the offense) than it does in cases where the defendant goes to trial (where judges hear more details about the offense). Future research might analyze courts' sentencing decisions for evidence of such an effect.

In advance of the proposed future work, results of the current studies offer robust evidence that the force of mitigating evidence depends on its timing. Laypersons assigned less severe punishment to perpetrators who experienced negative background circumstances as opposed to neutral or positive circumstances *only* when information about those circumstances was presented *before* information about their transgression. Continued research on this intriguing effect will help to further define its full scope, its boundaries, and its societal implications.

References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology: Attitudes and Social Cognition*, *63*, 368–378. doi: 10.1037/0022-3514.63.3.368

Appelbaum, P. S., & Scurich, N. (2014). Impact of behavioral genetic evidence on the adjudication of criminal behavior. *The Journal of the American Academy of Psychiatry and the Law*, *42*, 91–100.

Barnett, M. E., Brodsky, S. L., & Davis, C. M. (2004). When mitigation evidence makes a difference: Effects of psychological mitigating evidence on sentencing decisions in capital trials. *Behavioral Sciences & the Law*, *22*, 751-770. doi: 10.1002/bsl.591

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, *43*, 800. doi: 10.3758/s13428-011-0081-0

Bell Holleran, L. L., Vaughan, T. J., & Vandiver, D. M. (2016). Juror decision-making in death penalty sentencing when presented with defendant's history of child abuse or neglect. *Behavioral Sciences & the Law, 34*, 742-766. doi: 10.1002/bsl.2271

Bilz, K. (2010). We don't want to hear it: Psychology, literature and the narrative model of judging. *University of Illinois Law Review*, 429-488.

Berryessa, C. M. (2016). Genetic essentialist biases, stigma, and lack of mitigating impact on punishment decisions. *Journal of Law and the Biosciences*, *3*, 359–364. doi: 10.1093/jlb/lsw023

Bilz, K., & Nadler, J. (2009). Law, psychology, and morality. In B. H. Ross (Series Ed.), D. M.

Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Vol. Eds.), *Psychology of learning and motivation*: *Vol. 50. Moral judgment and decision making* (pp. 101–132). Burlington, MA: Academic Press. doi: 10.1016/S0079-7421(08)00403-9

Bilz, K. (2010). We don't want to hear it: Psychology, literature and the narrative model of judging. *University of Illinois Law Review*, 429-488.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3-5. doi: 10.1037/14805-009

Carlsmith, K. M., & Sood, A. M. (2009). The fine line between interrogation and retribution. *Journal of Experimental Social Psychology, 45*, 191-196. doi: 10.1016/j.jesp.2008.08.025

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. doi: 10.1037/0033-2909.112.1.155

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*, 353–380. doi: 10.1016/j.cognition.2008.03.006

Cushman, F. (2015). Punishment in humans: From intuitions to institutions. *Philosophy Compass, 10*, 117–133. doi: 10.1111/phc3.12192

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition, 127*, 6–21. doi: 10.1016/j.cognition.2012.11.008

Darley, J. (2009). Morality in the law: The psychological foundations of citizens' desires to punish transgressions. *Annual Review of Law and Social Science, 5*, 1–23. doi: 10.1146/annurev.lawsocsci.4.110707.172335

Darley, J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and just deserts as motives for punishment. *Law and Human Behavior, 24,* 659-683. doi: 10.1023/A:1005552203727

Drolet, C. E., Hafer, C. L., & Heuer, L. (2016). The role of perceived deservingness in the toleration of human rights violations. *Social Justice Research*, *29*, 429-455. doi: 10.1007/s11211-016-0273-y

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. doi: 10.3758/BF03193146

Federal Death Penalty Act of 1994, Pub. L. 103-322, title VI, § 60002(a), Sept. 13, 1994, 108 Stat. 1959, codified as amended at 18 U.S.C. § 3592.

Feigenson, N. R. (1995). The rhetoric of torts: How advocates help jurors think about causation, reasonableness, and responsibility. *Hastings Law Journal, 47,* 61-160. doi: 10.1037/10358-005

Forbes, R. C., & Stellar, J. E. (2022). When the ones we love misbehave: Exploring moral processes within intimate bonds. *Journal of Personality and Social Psychology*, *122,* 16–33. doi:10.1037/pspa0000272

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2000). Inside the judicial mind. *Cornell Law Review, 86,* 777-829. doi: 10.2139/ssrn.257634

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*, 557-559. doi: 10.1038/nature06288

Heuer, L. (2005). What's Just about the Criminal Justice System-A Psychological Perspective. *JL & Pol'y, 13*, 209-228. doi: 10.1177/0146167299258009

Heuer, L., Blumenthal, E., Douglas, A., & Weinblatt, T. (1999). A deservingness approach to respect as a relationally based fairness judgment. *Personality and Social Psychology Bulletin*, *25*, 1279-1292.

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General, 128*, 3-31. doi: 10.1037/0096-3445.128.1.3

Kahan, D. M. (2015). Laws of cognition and the cognition of law. *Cognition, 135*, 56-60. doi: 10.1016/j.cognition.2014.11.025

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An analysis of data quality: Professional panels, student subject pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*, 141-155. doi: 10.1080/00913367.2016.1269304

Kelman, M. (2013). Intuitions. *Stanford Law Review*, *65*, 1291-1323.

Landy, J. F., & Uhlmann, E. L. (2018). Morality is personal. In K. Gray, J. Graham (Eds.), *Atlas of Moral Psychology* (pp. 121-132). New York: Guilford Publications.

Levin, T. (1967). Toward a More Enlightened Sentencing Procedure. *Nebraska Law Review*, *45*, 499-512. doi: 10.1037/e438152008-038

Lynch, J. M., Lane, J. D. Berryessa, C. M., & Rottman, J. (2019). How information about perpetrators' nature and nurture influences assessments of their character, mental states, and deserved punishment. *PloS One*. doi: 10.1371/journal.pone.0224093

Meixner, J. B. (2022). Modern sentencing mitigation. *Northwestern University Law Review*, *116*, 1-72 (forthcoming).

Nadler, J. (2012). Blaming as a social process: The influence of character and moral emotion on blame. *Law and Contemporary Problems, 75*, 1-31.

Nadler, J., & McDonnell, M. H. (2011). Moral character, motive, and the psychology of blame. *Cornell Law Review, 97,* 255-300.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, *51*, 242-258. doi: 10.1037/0022-3514.51.2.242

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 521-533. doi: 10.1037/0278-7393.14.3.521

Pennington, N., & Hastie, R. (1990). Practical implications of psychological research on juror and jury decision making. *Personality and Social Psychology Bulletin*, *16*, 90–105. doi: 10.1177/0146167290161007

Pennington, N., & Hastie, R. (1991). A cognitive theory of juror decision making: The story model. *Cardozo Law Review*, *13*, 519-557.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62*, 189-206. doi: 10.1037/0022-3514.62.2.189

Rachlinski, J. J., & Wistrich, A. J. (2017). Judging the judiciary by the numbers: Empirical research on judges. *Annual Review of Law and Social Science, 13*, 203-229. doi: 10.1146/annurev-lawsocsci-110615-085032

Robbins, P., & Litton, P. (2018). Crime, punishment, and causation: The effect of etiological information on the perception of moral agency. *Psychology, Public Policy, and Law*, *24*, 118–127. doi: 10.1037/law0000146

Robinson, P. H., & Darley, J. M. (1995). *Justice, liability, and blame: Community views and the criminal law*. Boulder, CO: Westview. doi: 10.4324/9780429039812-7

Robinson, P. H., Jackowitz, S. E., & Bartels, D. M. (2012). Extralegal punishment factors: A study of forgiveness, hardship, good deeds, apology, remorse, and other such discretionary factors in assessing criminal punishment. *Vanderbilt Law Review, 65,* 737-826.

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, *10,* 72–81. doi: 10.1177/1745691614556679

Warriner, A.B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45,* 1191-1207.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of personality and social psychology*, *47*, 237-252. doi: 10.1037/0022-3514.47.2.237

Young, L., & Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. *Social and Personality Psychology Compass, 7*, 585–604. doi: 10.1111/spc3.12044

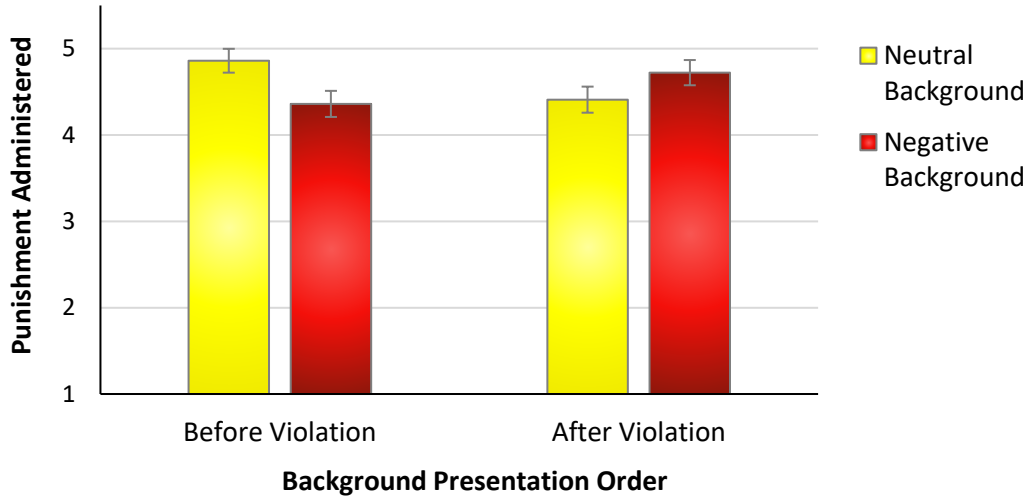

Figure 1. Study 1: Punishment decisions based on the nature of the perpetrator's background (neutral vs. negative) and the order in which the background information was presented (before vs. after the violation). Higher values on the y-axis indicate greater punishment. Error bars represent +/- 1 standard error of the mean.

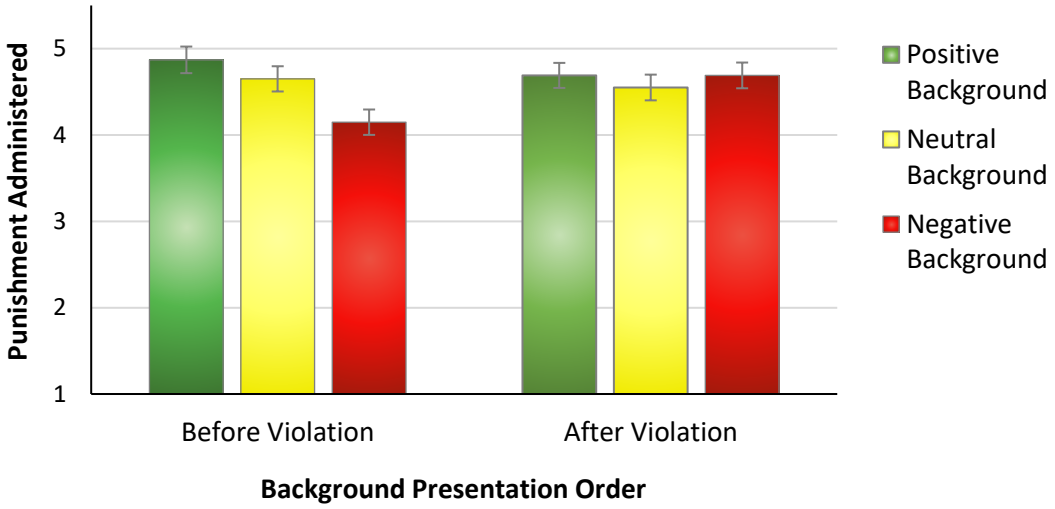

Figure 2. Study 2: Punishment decisions based on the nature of the perpetrator's background (positive, neutral, or negative) and the order in which the background information was presented (before vs. after the violation). Higher values on the y-axis indicate greater punishment. Error bars represent +/- 1 standard error of the mean.