Journal of Experimental Social Psychology 95 (2021) 104156

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Case ReportHow relationships bias moral reasoning: Neural and self-report evidence☆

Martha K. Berg *, Shinobu Kitayama, Ethan Kross

University of Michigan, United States of America

A R T I C L E I N F O

Keywords:
Morality
Interpersonal relationships
Expectancy violations
P300

A B S T R A C T

Laws govern society, regulating people’s behavior to create social harmony. Yet recent research indicates that when laws are broken by people we know and love, we consistently fail to report their crimes. Here we identify an expectancy-based cognitive mechanism that underlies this phenomenon and illustrate how it interacts with people’s motivations to predict their intentions to report crimes. Using a combination of self-report and brain (ERP) measures, we demonstrate that although witnessing any crime violates people’s expectations, expectancy violations are stronger when close (vs. distant) others commit crimes. We further employ an experimental-causal-chain design to show that people resolve their expectancy violations in *diametrically opposed* ways depending on their relationship to the transgressor. When close others commit crimes, people focus more on the individual (vs. the crime), which leads them to protect the transgressor. However, the reverse is true for distant others, which leads them to punish the transgressor. These findings highlight the sensitivity of early attentional processes to information about close relationships. They further demonstrate how these processes interact with motivation to shape moral decisions. Together, they help explain why people stubbornly protect close others, even in the face of severe crimes.

> *“When someone I knew, someone I had loved as a brother, was accused, I did something inexcusable. I publicly spoke up in his defense.”*
>
> –Lena Dunham (2018).

Consider the following: While at an electronics store with your best friend, you notice her grab an iPad and leave the store without paying. A few minutes later, a police officer approaches you and asks whether you saw your best friend steal something. What would you say?

What makes this scenario vexing is that it pits two fundamental drives against one another: protecting those we love (Aron, Aron, Tudor, & Nelson, 1991) versus abiding by universal rules governing society (Hofmann, Brandt, Wisneski, Rockenbach, & Skitka, 2018). Although research in moral psychology has grown exponentially over the past two decades, remarkably little research has examined what happens when these two drives collide. Instead, the majority of research on moral reasoning has focused on how people react to moral violations involving strangers (Bauman, McGraw, Bartels, & Warren, 2014; Bloom, 2011; Bostyn, Sevenhant, & Roets, 2018).

To address this gap, Weidman, Sowden, Berg, and Kross (2020) demonstrated that when people are faced with the choice of whether to protect or report an immoral actor, they are more likely to protect close (vs. distant) others. This effect increased with the severity of the crime people observed, was evident across multiple domains of moral transgression, and was uninfluenced by a variety of individual differences (for complementary evidence, see Hofmann et al., 2018; Waytz, Dungan, & Young, 2013).

Despite the strength and consistency of this phenomenon, two important questions remain. First, how does early cognitive processing shape these decisions? A growing body of evidence suggests that early attentional processes help shape moral decisions (e.g., Dubljević, Sattler, & Racine, 2018; Greene, 2017; Gui, Gan, & Liu, 2016; Haidt, 2001). Yet no research has examined how such processes are influenced by people’s relationships with those who commit crimes. By doing so, we aim to test how deeply the tendency to protect close others is rooted, which has important implications for theory and for future interventions. Second, how are these decisions influenced by the information people consider? Weidman et al. (2020) showed that people consider the harm that would come to a transgressor more strongly when a close (vs. distant) other commits a crime. However, no causal evidence exists linking these cognitive patterns to people’s decisions. Here, we address both questions to deepen our understanding of how relationships shape moral judgment.

☆ This paper has been recommended for acceptance by Dr Vanessa Bohns.

* Corresponding author at: Department of Psychology, University of Michigan, 530 Church St., Ann Arbor, MI 48109, United States of America.
E-mail address: bergmk@umich.edu (M.K. Berg).

https://doi.org/10.1016/j.jesp.2021.104156
Received 27 October 2020; Received in revised form 20 April 2021; Accepted 21 April 2021

1. Expectancy violations

A foundational assumption of modern psychology is that people develop cognitive representations early during development (Derryberry & Reed, 1996), which they use to form expectations about the world (Griffiths & Tenenbaum, 2011; Taylor & Crocker, 1981). These expectations help conserve effort as people navigate the world; only when expectations are violated are resources deployed to resolve the violation (Burgoon, 1993; Fiske & Taylor, 1991).

In the case of observing immoral behavior, we began with the assumption that all immoral acts violate people's expectations to some degree. This assumption is based on research indicating that people perceive others as moral (Dunning, Anderson, Schlösser, Ehlebracht, & Fetchenhauer, 2014) and expect them to behave accordingly (Reeder & Brewer, 1979). Therefore, any immoral behavior, regardless of one's relationship to the actor, should violate people's expectations. We substantiated this assumption in a series of self-report pretests (see Supplementary Materials).

Importantly, people's relationships to the person committing a crime may influence *the extent to which* their expectations are violated. Specifically, immoral behavior should be particularly unexpected when it comes from a close other (i.e., a main effect of closeness), given that close others overlap significantly with the self (Aron et al., 1991) and that people believe themselves to be less unethical than others (Klein & Epley, 2016). Thus, by extension, people should consider close others to be particularly unlikely to act immorally. Furthermore, close others' immoral behavior might be especially unexpected when they commit a highly severe act (i.e., an interaction between closeness and severity), which is less likely to occur than a low-severity act (i.e., a main effect of severity; Morgan & Oudekerk, 2019) and would be most inconsistent with people's positive beliefs about close others. Our first two goals were (1) to test these hypotheses, and (2) to examine how expectancy violations predict people's decisions to protect close vs. distant others.

2. Attentional deployment

Expectancy violations trigger a deployment of attentional resources that work toward resolving the violation (Burgoon, 1993; Fiske & Taylor, 1991). How people ultimately resolve the violation, however, is influenced by their motivations (Carver & Scheier, 1982). By way of analogy, attention is like the fuel that drives a car forward—yet whether the car goes to the right or the left is dependent on where the driver wants to go (how she turns the steering wheel).

Prior research indicates that people's relationships motivate them to reason in unique ways. Specifically, when a close other commits a crime, people are more likely to consider the harm that might come to the perpetrator, whereas when a distant other commits a crime, they are more likely to focus on the harm that the crime inflicts on society (Weidman et al., 2020). These findings suggest that people's goals (i.e., how they want to steer the metaphorical car) differ for close vs. distant others, which should shape the decisions they reach (i.e., where the car goes). Therefore, we expected people to resolve expectancy violations triggered by observing close (vs. distant) others' crimes in fundamentally different ways (see Fig. 1 for conceptual model). An expectancy violation about a close other should deploy attention toward the perpetrator, motivating a decision to protect them from consequences. In contrast, an expectancy violation about a distant other should deploy attention toward the crime itself, motivating a decision to report the perpetrator to the police. Although prior research has provided correlational evidence showing that relationships divert attention in a manner that is consistent with what we describe here, no study has provided causal evidence supporting this link. Therefore, our final goal was to provide causal evidence for the link between people's motivated patterns of attention and their decisions.

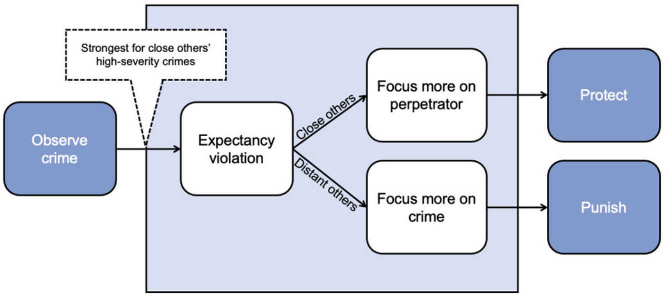

Fig. 1. Conceptual model. We expect people's prior experiences to shape the extent to which transgressions are unexpected, such that close others' crimes, and particularly their severe crimes, are most unexpected. Once an expectancy violation is triggered, we expect people's motivations to shape what information they attend to. For close others, people should be motivated to focus on the perpetrator, which should in turn make them more likely to protect the perpetrator. For distant others, people should be motivated to focus on the transgression, which should in turn make them more likely to punish the perpetrator.

3. Research overview

Three experiments examined how people's relationships shape their expectancy violations, their attentional focus once a violation has been triggered, and their decisions to protect a transgressor, across high- and low-severity moral transgressions. In Experiment 1, we examined people's brain activity as they imagined witnessing close and distant others committing immoral acts. We investigated how relational closeness influences expectancy violations, and how these in turn predict people's decisions to protect the transgressor. To do this, we used P300, an established electrocortical marker that is sensitive to violations of frequency-based expectations (Johnson, 1988), which allowed us to unobtrusively assess cognitive processes as they occurred, with millisecond precision (for complementary evidence at the self-report level, see Supplementary Materials). In Experiments 2a and 2b, we used self-report methods to examine how closeness to a transgressor influences people's attentional focus when they consider immoral acts (namely, whether people focus on the perpetrator vs. the crime), and how it predicts people's decisions. In all three experiments, we report all measures, manipulations and exclusions, and no additional participants were recruited after initial data analysis.

4. Experiment 1

Experiment 1 used event-related potentials (ERPs) to examine the role that expectancies play in people's responses to immoral behavior. Participants read about immoral acts that varied in (a) severity and (b) whether they were committed by a close versus distant other. We indexed expectancy violations using the P300, an ERP component that reflects the early (i.e., within 300 ms) deployment of cognitive resources to stimuli that are encoded as improbable and important (Johnson, 1988).

4.1. Method

4.1.1. Participants

The sample consisted of 59 European American, right-handed undergraduates with normal or corrected vision (for demographics, see Table S1). Three participants were excluded on a priori grounds— one for a past concussion and two for excessive noise during the task (>50% of trials rejected through standard processing).

Prior work indicates that in a within-subject, two-condition study, inter-participant P300 variability becomes stable with 5–7 participants (Yano, Suwazono, Arao, Yasunaga, & Oishi, 2019). We aimed for a substantially larger sample size given our 2 × 2 within-subject design. Therefore, we recruited as many subjects as possible during one academic

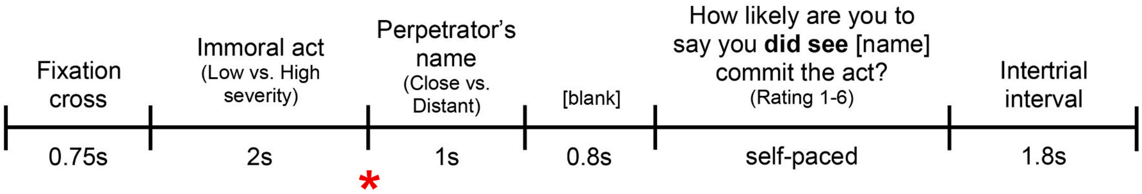

Fig. 2. Experiment 1 trial overview. For each trial, a brief description of an immoral act of either high or low severity was followed by the name of the perpetrator, who was either a close relation or a distant acquaintance. Subjects were then told to imagine that a police officer approached them and asked whether they had seen any suspicious activity. Finally, they rated the likelihood with which they would report the actions they witnessed. ERPs were time-locked to the presentation of the perpetrator's name (as indicated by the red asterisk). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

year.

4.1.2. Procedure

Participants first generated the names of three close and three distant others. To assist with name generation, participants were shown a diagram of concentric circles around the self, with close friends near the center and distant acquaintances further out. Participants were asked to briefly describe the relationship they shared with each person they nominated.

Next, participants completed 120 trials of a previously-developed paradigm (Weidman et al., 2020). On each trial, they imagined witnessing a high- or low-severity immoral act committed by a close or distant other, and they were asked to indicate how likely they would be to report what they had witnessed to a police officer using a 1 (*very unlikely*) to 6 (*very likely*) scale ($M = 3.33$, $SD = 1.97$; see Fig. 2 for trial overview and Table S2 for scenarios). Responses were reverse-coded so that higher values represented a greater likelihood of protecting the perpetrator.

4.1.3. EEG recording and processing

EEG was recorded using the BioSemi ActiveTwo System, with 32 scalp electrodes and 6 external electrodes configured to the 10–20 system. The data were digitized at 512 Hz, resampled offline at 256 Hz, and re-referenced to the average of the mastoids. Analysis was conducted using MATLAB software (MathWorks, 2017) with EEGLAB plugin (Delorme & Makeig, 2004) and ERPLAB Toolbox (Lopez-Calderon & Luck, 2014). Data were filtered offline using a low pass of 20 Hz and a high pass of 0.1 Hz. Artifacts were removed from continuous EEG data using independent components analysis (ICA). Decomposition of the independent components was performed using the 'runica' INFOMAX algorithm (Makeig, Jung, Bell, Ghahremani, & Sejnowski, 1997). Scalp electrodes deemed unsuitable for analysis were removed before performing ICA. The resulting components were visually inspected and artifactual components, which were primarily related to eye movements and muscle artifacts (McMenamin et al., 2010), were removed. Any removed channels were then interpolated spherically.

Next, data were time-locked to the presentation of the perpetrator's name and segmented into 1200-ms epochs, including a 200-ms prestimulus period for baseline correction. Bad electrodes, as determined through manual inspection of the data, were interpolated spherically. Data then underwent standard rejection procedures (Luck, 2014); rejected trials included those that exceeded $+/-$ 100 μV in a 200 ms moving window with a 50 ms step threshold, that fluctuated more than 30 μV between two sampling points, or that had little to no activity ($<$0.5 μV) throughout the trial.

Trial-level data were extracted from the Pz electrode site, where P300 is typically maximal (Picton, 1992). The maximum P300 peak of the grand-averaged waveform was visually identified, and a 60-ms time window was constructed around this peak to calculate the mean peak latency (334 ms). A 100-ms window around this latency (284–384 ms) was used to calculate mean amplitude for each trial. In a minority of trials (0.7%), P300 amplitude exceeded 3 standard deviations of the grand-mean (5.98 μV). Reported statistics correspond to the full dataset, but all model estimates were statistically equivalent with and without these outliers included.

4.1.4. Data analytic strategy

To test our hypotheses, we estimated three random effects models, each including random intercepts for each subject and moral scenario.[1] First, we tested the effect of closeness, severity, and their interaction on people's decisions. Second, we tested the effect of closeness, severity, and their interaction on expectancy violations (i.e., P300). Third, we tested how expectancy violations influenced the first model, by including grand-mean standardized P300 amplitude and all interactions as additional predictors. In each model, we probed key interactions by estimating simple effects at each level of the moderator.

4.2. Results

Self-report data from Experiment 1 directly replicated prior research: people were more likely to (a) protect close (vs. distant) others, and (b) protect perpetrators of low- (vs. high-) severity crimes. Moreover, the effect of closeness was greatest for high-severity crimes (see Table 1-A and Fig. 3-A).

Neural data revealed that close (vs. distant) others' actions elicited stronger expectancy violations, as indexed by P300 amplitude. High- (vs. low-) severity crimes elicited stronger expectancy violations, but this effect of severity was moderated by closeness (see Table 1-B and Fig. 3-B and C): Severity influenced expectancy violations for scenarios involving close others (simple effect: $b = 0.11$, 95% CI [0.04, 0.18], $p = .003$, $\eta p^2 = 0.05$), but not distant others (simple effect: $b = 0.01$, 95% CI [-0.06, 0.08], $p = .75$, $\eta p^2 < 0.001$). Together, these results suggest that relationships influence both (a) how unexpected immoral acts are and (b) how strongly severity information factors into people's early cognitive processing.

Next, we examined how expectancy violations indexed via P300 amplitude predicted participants' punishment decisions. These analyses revealed that P300 predicted participants' moral decisions about close and distant others in opposite directions. As Fig. 3-D illustrates, for transgressions committed by close friends, greater P300 amplitude predicted a greater likelihood of *protecting* the perpetrator. In contrast, for transgressions committed by distant acquaintance, greater P300 amplitude predicted a greater likelihood of *reporting* the perpetrator (P300 by closeness interaction: $b = 0.10$, 95% CI [0.05, 0.15], $p < .001$,

---

[1] All random effects models were fit using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) using maximum likelihood estimation. All p-values for model estimates were computed using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). Models reported in the main text are intercept-only models, which were selected a priori to account for between-subject and between-stimulus variation in average P300 amplitude and average behavioral intention. At the request of early readers, we also conducted post hoc analyses that included random slopes, which yielded consistent results (see Supplemental Materials). All R scripts will be made publicly available upon publication.

Table 1
Model estimates for key dependent variables in Experiment 1.

	Likelihood of reporting the perpetrator					P300 amplitude[c]				
Effect	Estimate	*SE*	95% CI		*p*	Estimate	*SE*	95% CI		*p*
			LL	*UL*				*LL*	*UL*	
Fixed effects										
Intercept	3.686	0.090	3.509	3.863	<0.001	0.002	0.045	−0.086	0.090	0.969
Closeness[a]	1.120[d]	0.025	1.071	1.169	<0.001	0.113[g]	0.024	0.067	0.159	<0.001
Severity[b]	−2.854[e]	0.106	−3.062	−2.646	<0.001	0.061[h]	0.029	0.005	0.118	0.037
Closeness x Severity	0.571[f]	0.050	0.472	0.669	<0.001	0.099[i]	0.047	0.007	0.191	0.035
Random effects										
By-crime variance	0.158		0.107	0.230		0.004		0.000	0.009	
By-subject variance	0.298		0.205	0.441		0.100		0.067	0.149	

Note. Number of subjects = 56, number of crimes = 60. (A) total $N = 6720$; (B) total $N = 6447$. CI = confidence interval; *LL* = lower limit; *UL* = upper limit. Sensitivity power analyses indicated that our design had 80% power to detect standardized beta coefficients of [d]0.019, [e]−0.088, [f]0.019, [g]0.033, [h]−0.045, and [i]0.034.

[a] −0.5 = distant, 0.5 = close.

[b] −0.5 = low severity, 0.5 = high severity.

[c] Grand-mean standardized.

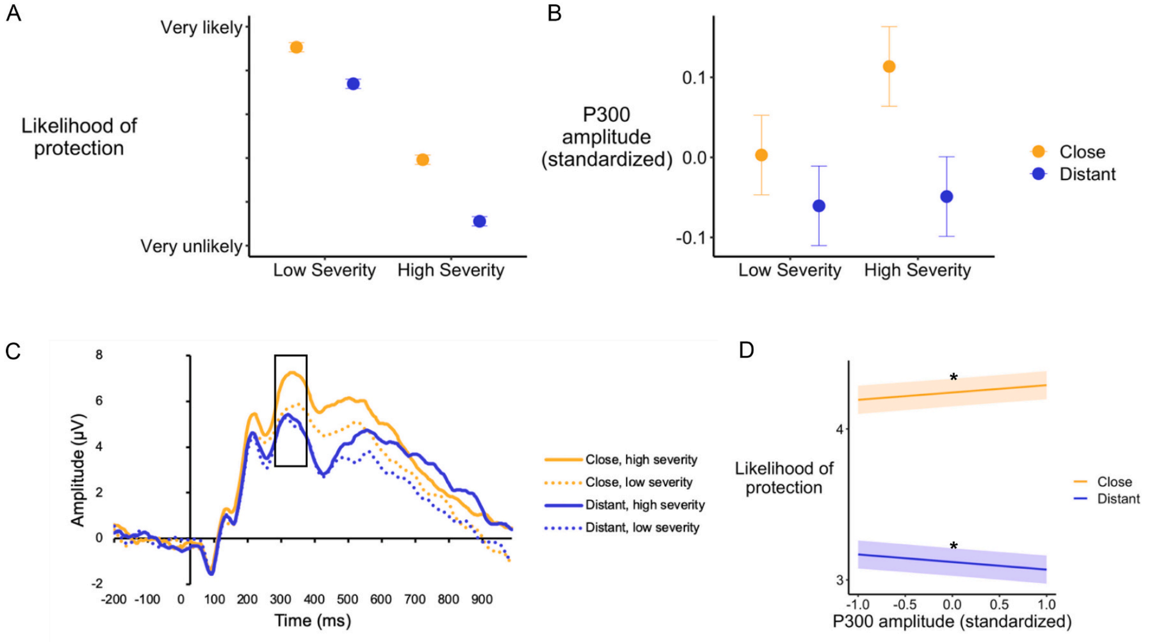

Fig. 3. Experiment 1 key findings. (A) Likelihood of protecting the perpetrator (reverse-scored from original question) based on closeness and severity. (B) Standardized P300 amplitude (in μV) based on closeness and severity. (C) Grand-averaged ERP waveform at Pz, baseline corrected and time-locked to the presentation of the perpetrator's name. Box indicates P300 time window from which mean amplitude was extracted. Note that the P300 deflection was observed in all conditions, indicating that all crimes were unexpected to some degree. (D) Standardized P300 amplitude (in μV) predicting participants' likelihood of protecting the perpetrator, depending on the relational closeness of the perpetrator. This effect was not significantly moderated by severity. All error bars and ribbons represent +/− 1 standard error. Asterisks indicate significant slopes.

$\eta p^2 = 0.002$; effect of P300 for close others: $b = 0.05$, 95% CI [0.01, 0.09], $p = .01$, $\eta p^2 = 0.04$; for distant others: $b = -0.05$, 95% CI [−0.09, −0.01], $p = .01$, $\eta p^2 = 0.05$).[2]

Importantly, expectancy violations predicted behavioral intentions in this way regardless of the severity of the crime participants observed. Specifically, severity did not moderate the interaction between P300 and closeness to predict participants' decisions (3-way interaction: $b = 0.06$, 95% CI [−0.04, 0.16], $p = .24$, $\eta p^2 < 0.001$), and results were consistent regardless of whether severity was included in the model.

To corroborate our neural findings, we turned to a series of self-report investigations (see Supplementary Materials for methods), which revealed that, consistent with the assumptions that underlie this work, people found all crimes unexpected to some degree. Furthermore, as revealed in the neural data, crimes committed by close (vs. distant) others, as well as high- (vs. low-) severity crimes, were particularly inconsistent with people's expectations (see Supplementary Experiments 1–4). Specifically, (a) people believed that close (vs. distant) others were more virtuous, and that high- (vs. low-) severity crimes were less frequent; and (b) people were more *confident* in their beliefs about close (vs. distant) others and high- (vs. low-) severity crimes. While we did not find convergent evidence supporting our secondary finding – namely, the interaction between closeness and severity – we speculate that this is due to the lower sensitivity of self-report measures, relative to the neural measures we used in Experiment 1.

In summary, Experiment 1 yielded three key findings. First, we

---

[2] A sensitivity power analysis indicated that our design had 80% power to detect a standardized beta coefficient of 0.018 for the interaction of P300 amplitude by closeness.

directly replicated prior work showing how closeness and severity predict decisions. Second, we found that relationships influence early (i.e., within 300 ms) cognitive processing; specifically, witnessing a close (vs. distant) others' moral transgression elicits a stronger expectancy violation. Third, we showed that relationships additionally influence *how* expectancy violations are resolved, once they are triggered.

It is noteworthy that although severity predicted expectancy violations, once the violation was triggered, its resolution was not further modulated by the severity of the original crime. Instead, the way people resolve these violations is determined only by their relationship to the transgressor, which we suspect is due to the underlying motivations people hold toward close vs. distant others—an idea that we examine directly in Experiments 2a and 2b.

5. Experiments 2a–b

In Experiment 1, we found that attentional deployment predicted behavioral intentions in opposite directions, depending on whether the expectancy violation involved a close or distant other. Our next goal was to examine the psychological mechanism that links expectancy violations to moral decisions. We predicted that when violations were the result of a close (vs. distant) other's moral transgression, they would deploy attention toward the person (vs. the transgression itself), which would in turn predict a more lenient decision. Previous work suggests that people consider different factors when making decisions about close (vs. distant) others, including thinking about harm to the perpetrator (vs. harm to others in society; Weidman et al., 2020), but no causal link has yet been established between attentional focus and moral decision.

Therefore, in Experiments 2a and 2b, we adopted an experimental-causal-chain approach (Spencer, Zanna, & Fong, 2005) to address this gap. Experiment 2a manipulated closeness and measured people's attentional focus and behavioral intentions; Experiment 2b manipulated attentional focus and assessed people's behavioral intentions.

6. Experiment 2a

6.1. Method

6.1.1. Participants

We recruited 400 participants through Prolific Academic (for demographics, see Table S1). We excluded one subject who indicated that the data they had provided was not reliable and/or valid.[3]

6.1.2. Procedure

The design for this study was identical to Experiment 1 with two exceptions. First, we used a between-subjects design, in which participants imagined witnessing one immoral act committed by a close or distant other, to probe people's decision processes without contaminating subsequent trials. The immoral act was either blackmail (high severity) or illegally downloading music (low severity), two scenarios that were used in Experiment 1 as well as in previous work (Weidman et al., 2020). Second, after deciding whether to report the perpetrator ($M = 2.42$, $SD = 1.71$), participants rated the extent to which they used information about (a) the person who committed the crime ($M = 4.43$, $SD = 2.30$) and (b) the immoral act itself ($M = 4.47$, $SD = 2.24$) when making their decisions, using a 1 (*not at all*) to 7 (*extremely*) scale.

6.1.3. Data analytic strategy

To test our hypotheses, we estimated three linear models. First, as in Experiment 1, we tested the effects of closeness, severity, and their interaction on people's decisions. Second, we examined how closeness influenced people's attentional focus by estimating a model predicting the extent to which people considered information (from *not at all* to *extremely*), based on the type of information (person or crime), relational closeness, severity, and all interactions. Third, we tested how attentional focus influenced the first model, by adding to this model the type of information (person or crime), the extent to which people considered that information, and all interactions. In each model, we probed key interactions by estimating simple effects at each level of the moderator.

Finally, as a supplementary analysis, we tested a mediational model using the lavaan package in R (Rosseel, 2012), with relational closeness as the independent variable and likelihood of protecting the perpetrator as the dependent variable. For the mediator, we computed an attentional focus difference score (person – crime).

6.2. Results

Replicating Experiment 1, participants were more likely to protect those who committed low- (vs. high-) severity crimes, and more likely to protect close (vs. distant) others, especially for severe transgressions (main effect of closeness: $b = 0.96$, 95% CI $= [0.69, 1.23]$, $p < .001$, $d = 0.56$; main effect of severity: $b = -1.80$, 95% CI $= [-2.07, -1.54]$, $p < .001$, $d = -1.22$; closeness by severity interaction: $b = 0.99$, 95% CI $= [0.46, 1.53]$, $p < .001$; $d = 0.29$).

Critically, participants focused on different information when making their decisions (closeness by information type interaction: $b = 2.25$, 95% CI $= [1.65, 2.85]$, $p < .001$; $d = 0.49$; see Fig. 4).[4] People were more likely to focus on information about the person (vs. the crime) for crimes committed by close others (effect of person vs. crime in close condition: $b = 1.06$, 95% CI $= [0.64, 1.49]$, $p < .001$, $d = 0.49$). In contrast, they were more likely to focus on the crime (vs. the person) for crimes committed by distant others (effect of person vs. crime in distant condition: $b = -1.19$, 95% CI $= [-1.62, -0.76]$, $p < .001$, $d = -0.54$).

As in Experiment 1, closeness predicted attentional focus in this way regardless of the severity of the crime participants observed. Specifically, severity did not moderate the interaction between closeness and information type to predict participants' weighting of the information (3-way interaction: $b = 1.08$, 95% CI $= [-0.12, 2.28]$, $p = .08$, $d = 0.12$), and results were identical whether or not severity was controlled for in the model. Thus, once again, people's closeness to a perpetrator influenced their attention (on the person vs. the crime), but the severity of the crime did not.

Moreover, consistent with our interpretation of the Experiment 1 findings, the more people focused on the perpetrator, the more likely they were to protect them. In contrast, the more people focused on the crime, the less likely they were to protect the perpetrator (information type x weight interaction: $b = 0.37$, 95% CI $[0.27, 0.48]$, $p < .001$, $\eta p^2 = 0.04$; simple slope for person: $b = 0.09$, 95% CI $[0.02, 0.16]$, $p = .01$; for immoral act: $b = -0.28$, 95% CI $[-0.36, -0.21]$, $p < .001$). Cross-sectional mediation analyses revealed that the extent to which people considered the person (vs. the crime) significantly mediated the relation between closeness and people's decisions (indirect effect: $b = 0.30$, 95% CI $[0.16, 0.48]$, $p < .001$; see Supplementary Materials; see Experiment 2b for corresponding causal analyses).

In summary, Experiment 2a replicated the key findings from Experiment 1 and demonstrated that when people think about transgressions, they spontaneously focus on different information depending on their relationship to the perpetrator. This, in turn, predicts moral decisions: whereas focusing on the person predicts protecting the perpetrator, focusing on the crime itself predicts punishment. This constitutes preliminary evidence for a mediational framework, in which

---

[3] In Experiments 2 and 3, at the end of the study session, participants were shown the following question: "It is very important for us to have reliable and valid data. Would you recommend that we use your responses to this survey as part of our study?" In both experiments, we excluded any participants who answered "No" to this question.

[4] A sensitivity power analysis indicated that our design had 80% power to detect a standardized beta coefficient of 0.142.

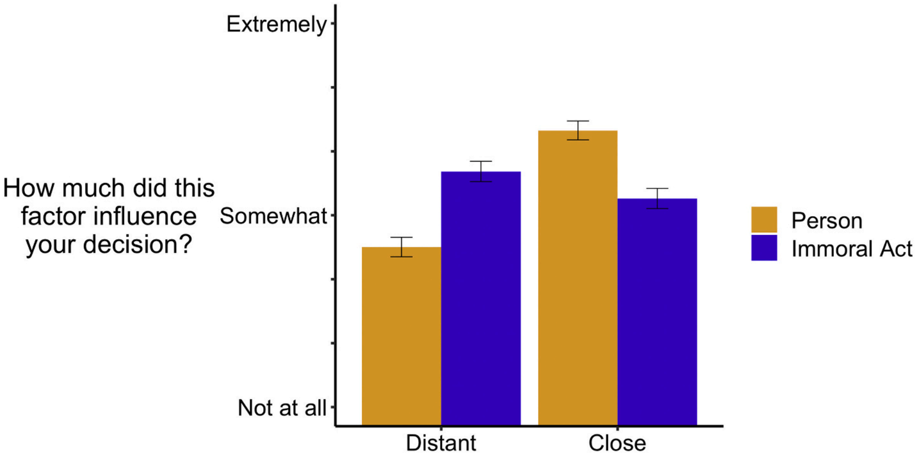

Fig. 4. Self-reported weighting of each factor—person and immoral act—by relational closeness. Error bars represent +/− 1 standard error.

witnessing these unexpected events leads to a motivated deployment of attention, which in turn predicts behavioral intentions. Nevertheless, as we did not manipulate which information participants focused on, the latter path of this proposed framework is based on correlational evidence. Thus, we designed Experiment 2b to fill this gap.

7. Experiment 2b

To provide causal evidence that focusing on the perpetrator vs. the crime influences people's decisions, we asked participants to read a scenario in which they observed a close or distant other commit a severe crime. We focused on high-severity crimes for two reasons. First, our first two experiments indicated that the effect of closeness on attentional focus is not dependent on the severity of the crime. Second, high-severity crimes represent the strongest test of our hypothesis, given that the stakes of the moral decision are highest.

Before participants made their decisions, we asked them to focus on either the person involved in the scenario (their close or distant other) or the crime (namely, blackmailing a person for money). Consistent with Experiment 2a findings, we expected that when people focused on the perpetrator (vs. the crime), they would be more likely to protect the perpetrator, and we expected this effect to be irrespective of relational closeness.

7.1. Method

7.1.1. Participants

We recruited 799 participants through Prolific Academic (for demographics, see Table S1).[5] We excluded 10 subjects who indicated that the data they had provided was not reliable and/or valid. We also excluded 336 subjects who spent 12 s or less reading the moral scenario (although it is noteworthy that including these participants in our analyses did not alter the significance of our results; see Supplementary Materials). We adopted the latter a priori exclusion criteria after piloting the materials for Experiment 2b. During piloting, we discovered that a disproportionate number of subjects completed the study without spending adequate time reading the moral scenarios. Therefore, we selected a threshold of 12 s, which we preregistered before data collection (https://aspredicted.org/blind.php?x=a2za7t).[6]

7.1.2. Procedure

The procedure was identical to that of Experiment 2a with two exceptions. First, the two manipulated factors were closeness and information type (person vs. crime); all participants saw the high-severity scenario from Experiment 2a (blackmail). Second, before answering the police officer, participants spent 30 s thinking about the assigned information type: either *the person involved in this scenario – [name]* or *the crime that was committed – blackmailing a person for money*. Afterward, participants indicated how likely they were to report the perpetrator to the police, from 1 *(very unlikely)* to 6 *(very likely*; $M = 3.33$, $SD = 1.74$*)*.

7.1.3. Data analytic strategy

To test our hypothesis, we estimated a linear regression predicting people's intention to protect the perpetrator from closeness, attentional focus (person vs. crime), and their interaction.

7.2. Results

Consistent with previous studies, participants were more likely to protect close versus distant others ($b = 1.75$, 95% CI [1.46, 2.03], $p < .001$, $d = 1.16$).

As expected, when people focused on the person (vs. the crime) before making their decisions, they were more likely to protect the perpetrator ($b = 0.34$, 95% CI [0.06, 0.61], $p = .02$, $d = 0.17$; see Fig. 5).[7] As in Experiment 2a, this effect was independent of people's relationship to the perpetrator (closeness x information type interaction: $b = 0.16$, 95% CI [−0.39, 0.72], $p = .57$, $d = 0.05$).

In summary, Experiment 2b provided causal evidence for the final link in our proposed mediational pathway. Specifically, we show that when participants are experimentally induced to focus on the person committing the crime (which people tend to spontaneously do for close others, as shown in Experiment 2a), they are more likely to protect the perpetrator. However, when participants are experimentally induced to focus on the crime itself (which people tend to spontaneously do for distant others), they are more likely to punish the perpetrator.

7.3. Discussion

Kidnapping. Armed robbery. Blackmail.

At first blush, when we consider how to respond to witnessing a person commit such crimes, the answer seems obvious: we should report them. Yet when we imagine that the perpetrator is someone we love,

---

[5] Given that only approximately 50% of our pilot sample met our exclusion criteria, we oversampled our target sample size by 100%.

[6] For consistency, we reanalyzed our findings from Experiment 2a using the 12-s reading exclusion criterion. All findings were identical with and without these exclusions applied (see Supplementary Materials).

[7] A sensitivity power analysis indicated that our design had 80% power to detect a standardized beta coefficient of 0.117.

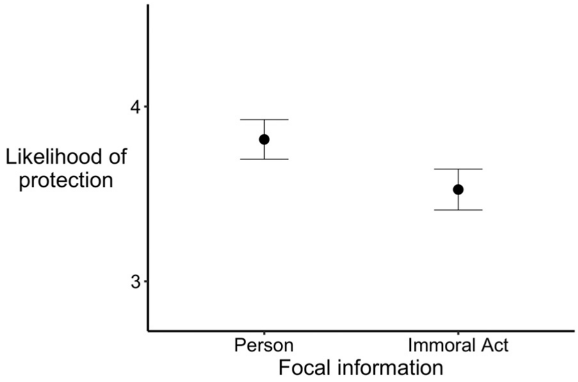

Fig. 5. Likelihood of protection predicted by manipulated focal information—person vs. immoral act. Error bars represent +/− 1 standard error.

both the real stories of people like Lena Dunham and previous empirical studies indicate that the opposite answer—to protect them—seems equally obvious. What explains this contradiction? Here, we begin to address this question by demonstrating how our relationships shape fundamental cognitive processes, including both the early orientation of attention to an unexpected event and the motivation-guided resolution of this violation.

Crimes, particularly those that are committed by close others, violate people's expectations, as evidenced by people's early neural responses. In turn, these violations lead to a resolution that is consistent with people's goals. Because people are driven by different motivations when considering close versus distant others, this process yields different decisions depending on one's relationship to the perpetrator.

The more people consider the perpetrator of the act in making their decisions (which is more likely when people consider close others' crimes), the more likely they are to *protect* that person. In contrast, the more people consider the immoral act in making their decisions (which is more likely when people consider distant others' crimes), the more likely they are to *report* the act.

It is important to note that even though people consider close (vs. distant) others to be more virtuous (see Supplemental Experiment 2), once someone commits a crime, people see it as equally immoral regardless of their relationship with a transgressor (Weidman et al., 2020, Study 2). However, as we show throughout the present work, despite this seeming impartiality in judging crimes, people make different decisions in response to close and distant others' behaviors. Thus, it is possible that whereas people's judgments of an act are colored by a desire to appear impartial, their behavioral intentions reveal their motivations to protect the people they love.

Of course, whether or not one *should* report someone who commits a crime is a separate question; the current findings do not speak to this issue. That said, one could imagine certain situations in which it would be beneficial to identify ways of reducing this bias, especially given that it is strongest when crimes are severe and therefore may carry grave consequences. In this vein, the current findings highlight multiple potential points of intervention to work toward these goals. People's expectations about a transgressor could be manipulated, for instance, by providing additional information from which to base expectations. People's attentional resources could also be manipulated, for example by increasing cognitive load, which would reduce their ability to deploy attention in motivated ways, thereby potentially mitigating the sway of relationships on moral decisions. Finally, interventions could be designed to reshape the motivations underlying the resolution of expectancy violations, which would lead to different behavioral intentions.

Notably, neural and self-report methods converged to indicate that the severity of the crime did not influence the extent to which people engaged in motivated cognition about immoral acts. This finding underscores the power of relationships to shape people's moral decisions. Once an expectancy violation is triggered, it is not the severity of the act but one's *relationship with the transgressor* that drives how the expectancy violation is resolved. However, it is possible that more vividly experiencing a severe immoral act, such as seeing it unfold in person rather than imagining it as a hypothetical scenario, would push people to direct more attention to the crime, even for close others. Our model suggests that this, in turn, would predict less protection of close others. Future research should examine this possibility.

Another important question for future examination is whether these findings prevail in other cultural contexts. While the information-processing principles motivating this account are believed to be invariant across cultures, they may interact with cultural values. For example, loyalty to friends or immediate kin (as in our close other condition) may be weaker in collectivist contexts such as East Asian societies or military settings. Instead, loyalty may be devoted primarily to higher social units such as extended family or country. In this case, loyalty would be aligned with the goal of condemning moral harm, even when the crime involves a close other, which could result in a weaker bias toward protecting close others in these contexts.

Additionally, future research should explore how other relational dynamics influence the processes we identify here. For example, how might these cognitive processes shift when one is the *victim* of a close (vs. distant) other's immoral act? Understanding how people respond to their *own* victimization by a close other would not only contribute to a more nuanced understanding of the cognitive processes we investigate here, but it would also shed light on the pervasive underreporting that has been documented, for example, among survivors of sexual assault (Morgan & Oudekerk, 2019), a crime that is frequently committed by a close other (Black et al., 2011).

In sum, this work provides neural and self-report evidence for an expectancy-based mechanism underlying moral decisions about close others. The relationship-dependent association we reveal between neural responses and behavioral intentions helps to fill an important gap in the moral psychology literature, moving beyond a model of immoral strangers to put forward a more nuanced framework of how individuals respond to immoral actors they know and love.

Author contributions

M.K.B., S.K., and E.K. developed the study concept and designed and conducted the studies. M.K.B. analyzed the data and wrote the article, with assistance from E.K. Critical revisions were provided by S.K. All authors approved the article for submission.

Open practices

All data, code, and materials are available at the following link: https://osf.io/6au4z/?view_only=c89f8750f86341669f73401846aca1b6.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2021.104156.

References

Aron, A., Aron, E. N., Tudor, M., & Nelson, G. (1991). Close relationships as including other in the self. *Journal of Personality and Social Psychology, 60*(2), 241–253. https://doi.org/10.1037/0022-3514.60.2.241.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass, 8*(9), 536–554. https://doi.org/10.1111/spc3.12131.

Black, M. C., Basile, K. C., Breiding, M. J., Smith, S. G., Walters, M. L., Merrick, M. T., … Stevens, M. R. (2011). The national intimate partner and sexual violence survey: 2010 summary report. *Centers for Disease Control and Prevention, 124*.

Bloom, P. (2011). Family, community, trolley problems, and the crisis in moral psychology. *The Yale Review, 99*(2), 26–43. https://doi.org/10.1111/j.1467-9736.2011.00701.x.

Bostyn, D. H., Sevenhant, S., & Roets, A. (2018). Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychological Science, 29*(7), 1084–1093. https://doi.org/10.1177/0956797617752640.

Burgoon, J. K. (1993). Interpersonal expectations, expectancy violations, and emotional communication. *Journal of Language and Social Psychology, 12*(1–2), 30–48. https://doi.org/10.1177/0261927X93121003.

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin, 92*(1), 111–135. https://doi.org/10.1037/0033-2909.92.1.111.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Derryberry, D., & Reed, M. A. (1996). Regulatory processes and the development of cognitive representations. *Development and Psychopathology, 8*(1), 215–234. https://doi.org/10.1017/S0954579400007057.

Dubljević, V., Sattler, S., & Racine, E. (2018). Deciphering moral intuition: How agents, deeds, and consequences influence moral judgment. *PLoS One, 13*(10), Article e0204631. https://doi.org/10.1371/journal.pone.0204631.

Dunning, D., Anderson, J. E., Schlösser, T., Ehlebracht, D., & Fetchenhauer, D. (2014). Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology, 107*(1), 122–141. https://doi.org/10.1037/a0036673.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. McGraw-Hill.

Greene, J. D. (2017). The rat-a-gorical imperative: Moral intuition and the limits of affective learning. *Cognition, 167*, 66–77. https://doi.org/10.1016/j.cognition.2017.03.004.

Griffiths, T. L., & Tenenbaum, J. B. (2011). *Predicting the future as Bayesian inference: People combine prior knowledge with observations when estimating duration and extent*. Other University Web Domain. https://dspace.mit.edu/handle/1721.1/70990.

Gui, D.-Y., Gan, T., & Liu, C. (2016). Neural evidence for moral intuition and the temporal dynamics of interactions between emotional processes and moral cognition. *Social Neuroscience, 11*(4), 380–394. https://doi.org/10.1080/17470919.2015.1081401.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814.

Hofmann, W., Brandt, M. J., Wisneski, D. C., Rockenbach, B., & Skitka, L. J. (2018). Moral punishment in everyday life. *Personality and Social Psychology Bulletin, 44*(12), 1697–1711. https://doi.org/10.1177/0146167218775075.

Johnson, R. (1988). The amplitude of the P300 component of the event-related potential: Review and synthesis. *Advances in Psychophysiology, 3*, 69–137.

Klein, N., & Epley, N. (2016). Maybe holier, but definitely less evil, than you: Bounded self-righteousness in social judgment. *Journal of Personality and Social Psychology, 110*(5), 660–674. https://doi.org/10.1037/pspa0000050.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8. https://doi.org/10.3389/fnhum.2014.00213.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.

Makeig, S., Jung, T.-P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences, 94*(20), 10979–10984. https://doi.org/10.1073/pnas.94.20.10979.

MathWorks. (2017). *MATLAB 2017a*.

McMenamin, B. W., Shackman, A. J., Maxwell, J. S., Bachhuber, D. R. W., Koppenhaver, A. M., Greischar, L. L., & Davidson, R. J. (2010). Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. *NeuroImage, 49*(3), 2416–2432. https://doi.org/10.1016/j.neuroimage.2009.10.010.

Morgan, R. E., & Oudekerk, B. (2019). *Criminal victimization, 2018 (NCJ 253043). 37*. U. S. Department of Justice, Bureau of Justice Statistics.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology, 9*(4), 456–479. https://doi.org/10.1097/00004691-199210000-00002.

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review, 86*(1), 61–79. https://doi.org/10.1037/0033-295X.86.1.61.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(1), 1–36. https://doi.org/10.18637/jss.v048.i02.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89*(6), 845–851. https://doi.org/10.1037/0022-3514.89.6.845.

Taylor, S. E., & Crocker, J. (1981). Schematic bases of social information processing. In E. T. Higgins, C. P. Herman, & M. Zanna (Eds.), *Social cognition*. Lawrence Erlbaum Associates.

Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower's dilemma and the fairness–loyalty tradeoff. *Journal of Experimental Social Psychology, 49*(6), 1027–1033. https://doi.org/10.1016/j.jesp.2013.07.002.

Weidman, A. C., Sowden, W. J., Berg, M. K., & Kross, E. (2020). Punish or protect? How close relationships shape responses to moral violations. *Personality and Social Psychology Bulletin, 46*(5), 693–708.

Yano, M., Suwazono, S., Arao, H., Yasunaga, D., & Oishi, H. (2019). Inter-participant variabilities and sample sizes in P300 and P600. *International Journal of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2019.03.010.