In press, *Journal of Personality and Social Psychology*

People Systematically Update Moral Judgments of Blame

Andrew E. Monroe[1]* and Bertram F. Malle[2]
[1]Appalachian State University, [2]Brown University

Abstract

Six experiments examine people's updating of blame judgments and test predictions developed from a socially-regulated blame perspective. According to this perspective, blame emerged in human history as a socially costly tool for regulating other's behavior. Because it is costly for both blamers and violators, blame is typically constrained by requirements for "warrant"—evidence that one's moral judgment is justified. This requirement motivates people to systematically process available causal and mental information surrounding a violation. That is, people are relatively calibrated and even-handed in utilizing evidence that either amplifies or mitigates blame. Such systematic processing should be particularly visible when people update their moral judgments. Using a novel experimental paradigm, we test two sets of predictions derived from the socially-regulated blame perspective and compare them with predictions from a motivated-blame perspective. Studies 1-4 demonstrate (across student, internet, and community samples) that moral perceivers systematically grade updated blame judgments in response to the strength of new causal and mental information, without anchoring on initial evaluations. Further, these studies reveal that perceivers update blame judgments symmetrically in response to exacerbating and mitigating information, inconsistent with motivated-blame predictions. Study 5 shows that graded and symmetric blame updating is robust under cognitive load. Lastly, Study 6 demonstrates that biases can emerge once the social requirement for warrant is relaxed—as in the case of judging outgroup members. We conclude that social constraints on blame judgments render the normal process of blame well calibrated to causal and mental information, and biases may appear when such constraints are absent.

Keywords: Moral Judgment, Blame, Intentionality, Mental States, Motivational Bias, Anchoring and Adjustment

---

*Correspondence to:
Andrew E. Monroe
Department of Psychology
Appalachian State University
222 Joyce Lawrence Lane
Boone, NC 28608 USA
E-mail: monroeae1@appstate.edu

People Systematically Update Moral Judgments of Blame

Questions about morality abound in human social life. One can hardly read a newspaper, watch TV, or converse with friends without encountering issues of blame, praise, moral responsibility, or moral character. Indeed, judgments of morality predominate people's enduring sense of self (Aquino & Reed, 2002; Strohminger & Nichols, 2015); perceptions of morality are central to social impressions of others (Goodwin, Piazza, & Rozin, 2014; Wojciszke, Bazinska, & Jaworski, 1998); and moral norms and taboos bind individuals together into (largely) cooperative communities (Bicchieri, 2006; Haidt, 2008; Wilson, 2010).

Among the various types of moral judgments that people render, judgments of blame carry particular social significance. Whereas moral judgments of permissibility, badness, or wrongness are directed at *behaviors* that violate moral standards (see Uhlmann, Pizarro, & Diermeier, 2015), judgments of blame single out the *person* who violated the standard. And when blame judgments are expressed socially, they impose costs on the violator through personal criticism and challenges to social standing. Because of blame's social nature, it carries substantial risks for blamers as well, such as retaliation by the criticized norm violator or reputational damage when an accusation turns out to be unfounded. Some theorists argue that, to minimize such costs and risks, acts of blaming are tightly regulated by norms of moral criticism (Bergmann, 1998; Coates & Tognazzini, 2012; Ingram, 2014; Malle, Guglielmo, & Monroe, 2014; Voiklis & Malle, 2017). These norms demand that when people blame others they are required to have *warrant*—evidence that one's moral judgment is fair and justified.

These inherent social constraints on blame lead to an intriguing possibility: The requirement for warrant and the potential social cost of blaming may motivate people to be relatively careful in attending to available blame-relevant information, including agents' intentionality and mental states, their causal contributions to an outcome, and even counterfactuals about the preventability of the outcome (Cushman, 2008; Gray, Young, & Waytz, 2012; Guglielmo, Monroe, & Malle, 2009; Malle, Guglielmo, & Monroe, 2012; Malle et al., 2014; Monroe & Malle, 2017; Shaver, 1985; Weiner, 1995; Young & Saxe, 2009). According to this perspective, blame not only regulates other people's behavior, but blame itself is *socially regulated,* motivating people to be systematic in their information processing toward blame.

The psychological literature, however, often paints a less optimistic view of moral cognition. A family of theories, which we collectively refer to as *motivated-blame* models, suggest that consideration of causal and mental information is secondary to and biased by early-emerging moral judgments and a general desire to blame.[1] On this view moral judgments quickly emerge in response to a norm violation, and people consider the details of the event (e.g., intentionality, reasons, controllability) later, often as a post-hoc rationalization of their judgments (Alicke, 2000; Haidt, 2001; Knobe, 2003; Pettit & Knobe, 2009; Tetlock et al., 2007). For example, Greene (2008) describes humans as "creatures who exhibit social and moral behavior that is driven largely by intuitive emotional responses and who are prone to rationalization of their behaviors" (pp. 62-63).

[1] The various models differ in how they label this early stage. Some refer to intuitive judgments (Jonathan Haidt, 2001), others refer to emotional responses (Greene, 2008), yet others refer to evaluative reactions (Mark D. Alicke, 2000), and some leave their character unanalyzed but label them genuinely moral (Pettit & Knobe, 2009). All of them share in common the suggestion that these early processes precede and often bias systematic consideration of mental and causal information.

Comparing the socially-regulated and motivated-blame perspectives has proved difficult (see Guglielmo, 2015). The two perspectives, while not mutually exclusive, do however make divergent predictions about how people update moral judgments. Updating refers to making a moral judgment and then learning new information (mitigating or exacerbating) that invites a revision of the initial judgment. For such moral updating situations, the socially-regulated blame perspective suggests that demands for warrant motivate perceivers to engage in relatively systematic processing of available causal and mental information, including even-handedly weighing mitigating and exacerbating information. By contrast, motivated-blame models suggest that a desire to blame motivates people to engage in biased judgment revisions, asymmetrically favoring information that confirms or exacerbates existing blame judgment over information that mitigates blame (Alicke, 2000; Ames & Fiske, 2013).

In the present studies we introduce a new experimental paradigm that models a moral updating situation. The process of updating person representations is well documented in the impression formation literature (Cone & Ferguson, 2015; Kammrath, Ames, & Scholer, 2007; Mende-Siedlecki, Cai, & Todorov, 2012; Rapp & Kendeou, 2007; Reeder & Brewer, 1979), but it has not been explored in the moral domain. In our paradigm (Monroe & Malle, 2017), people (1) encounter a sparse description of an immoral event, (2) make an initial blame judgment of that event, (3) receive additional information about the perpetrator's mental states or causal contributions (that could mitigate or exacerbate blame), and (4) have the opportunity to register an updated blame judgment. This paradigm allows us to test fine-grained predictions about graded judgment updates that are systematically responsive to different types of information—predictions that fall out of the recently proposed Path Model of Blame (Malle et al., 2014). Additionally, changes from initial to updated blame judgments allow us to compare the Path Model's prediction of symmetric updating (perceivers are equally responsive to mitigating as to exacerbating information) with motivated-blame models' prediction of asymmetric updating (perceivers are more responsive to exacerbating than to mitigating information). Below we review the socially-regulated blame and motivated-blame perspectives in more detail and develop their predictions for moral updating.

Theoretical Background

The Socially-Regulated Blame Perspective

Theorists broadly agree that morality evolved to facilitate group life (Carnes, Lickel, & Janoff-Bulman, 2015; Haidt, 2007; Malle et al., 2012; Rai & Fiske, 2011). Indeed blame—as socially expressed disapproval—may be one of the oldest tools for human behavior regulation (Przepiorka & Berger, 2016; Voiklis & Malle, 2017) and is effective at enforcing cooperation (Guala, 2012). Yet, in order to accomplish its social-regulatory function blame must be publicly expressed, either to moral offenders or to others as gossip, and as such it imposes costs on the alleged offender (e.g., loss of face, status, or reputation) and also comes with risks for the blamer if the accusation is unfounded or the offender retaliates.

Indeed, the social context in which blame emerged highlights the potential costs of unfounded blame. In the 40-80,000 years before human settlements, humans lived in small nomadic bands where cooperation and maintaining relationships was critical to survival (Boehm, 2000; Knauft, 1991). In these bands, norm violations and people's responses to them were inherently transparent affairs (Silberbauer, 1982; Wilson, 1991). Thus, to keep these costs in

check, and to maintain fair treatment (Wallace, 1994), acts of blaming became regulated by social norms of moral criticism (Coates & Tognazzini, 2012; Malle et al., 2014).2

In this line, recent research demonstrates that people are strongly averse to overblaming (Kim, Voiklis, Cusimano, & Malle, 2015) and to unwarranted blame (Mikula, Petri, & Tanzer, 1990; Mikula, Scherer, & Athenstaedt, 1998). People react negatively when they feel unfairly blamed (MacCoun, 2005; Miller, 2001), and even preschool children are willing to correct an adult who unfairly punishes another agent for an accidental transgression (Chernyak & Sobel, 2016). Conversely, failures to blame and punish carry similar risks. For example, in June 2018, Aaron Persky—the judge who presided over the infamous Brock Turner sexual assault case—was recalled, largely due to public outrage over the perception that he failed to sufficiently punish Turner. Likewise, recent empirical work demonstrates that people who decide to forgive wrongdoers rather than punishing them are perceived as blameworthy and as having bad moral character (Gardner & Monroe, 2018).

Thus, moral judges must walk a fine line: over-blaming risks reactive aggression from targets; whereas, under-blaming risks being censured in turn. Appreciating the social context in which acts of blame occur suggests that moral perceivers are motivated to "get blame right" or, at a minimum to make judgments broadly perceived as fair. But what does "getting blame right" mean? It means grounding one's judgment in just the kind of evidence that community members routinely use in forming and checking blame judgments: information about the severity of harm, causality, intentionality, and mental states (Cushman, 2008; Lagnado & Channon, 2008; Malle, 1999; Malle et al., 2014; Reeder & Coovert, 1986; Young & Saxe, 2009).

This theoretical perspective suggests that people will flexibly revise blame judgments in response to new, morally-relevant information (e.g., intentionality, reasons, or outcome preventability), regardless of whether the information supports increasing or decreasing blame. On the face of it, these prediction appear to contradict well-documented findings on general information processing biases, especially the confirmation bias (Ditto & Lopez, 1992; Gilovich, 1983). However, recent research on non-moral person perception hints at conditions under which people readily update their representations of others' character (Mende-Siedlecki et al., 2012), namely when they identify new information that is diagnostic and meaningful as opposed to merely inconsistent with a previous impression (Cone & Ferguson, 2015; Cone, Mann, & Ferguson, 2017; Mende-Siedlecki, Baron, & Todorov, 2013). And because social demands put a premium on diagnostic information about norm violators, confirmation bias may well play a more limited role in interpersonal moral judgments of blame.

Indeed, this prediction is supported by research on accountability. When people are publicly accountable for their judgments they are more likely to overcome common cognitive biases (Tetlock, 1985), engage in flexible and systematic information processing (Scholten, van Knippenberg, Nijstad, & De Dreu, 2007; Tetlock, Skitka, & Boettger, 1989), and make more nuanced judgments about moral responsibility (Lerner, Goldberg, & Tetlock, 1998). However, the socially-regulated blame perspective goes one step further in suggesting that blame judgments will be nuanced and systematic not only under explicit accountability demands but whenever blame judgments are publicly expressed, including in the context of an experiment. As a result, blame should be generally less susceptible to confirmation bias than are other moral and

---

[2] We want to emphasize that these considerations of blame do not extend to mere judgments of badness, permissibility, or wrongness, which reflect a moral evaluation of a behavior and not a judgment of the whole person. For a discussion of this distinction see Malle et al. (2014, pp. 148-150) and Pizarro and Tannenbaum (2012).

nonmoral judgments. However, arguing that blame is socially regulated does not imply that people are perfectly systematic or calibrated in their judgments. Rather, the potential social costs of over- or under-blaming prompt people to attend to available information and to strive to adjust their initial judgment in light of it.

The Motivated-Blame Perspective

Virtually all perspectives on moral judgment agree that people respond to norm-violating events with rapid evaluation (Luo et al., 2006; Van Berkum, Holleman, Nieuwland, Otten, & Murre, 2009) that activates further information processing (Mikhail, 2007). According to the motivated-blame perspective, more specifically, early-emerging moral evaluations and a desire to blame bias subsequent information processing of causal and mental-state information in favor of confirming or strengthening blame (Alicke, 1992; Ditto, Pizarro, & Tannenbaum, 2009; Mazzocco, Alicke, & Davis, 2004; Nadelhoffer, 2006; Nadler, 2012). For example, Alicke and colleagues argue that, "Negative evaluations or spontaneous reactions lead to the hypothesis that the source of the evaluations is blameworthy, and to an active desire to blame that source. This desire, in turn, leads observers to interpret the available evidence in a way that supports their blame hypothesis" (Alicke, Rose, & Bloom, 2011, p. 675). Similarly, Ditto and colleagues propose that "Moral judgments are most typically top-down affairs, with the individual generating moral arguments with intuitions about the "correct" moral conclusion already firmly in place" (2009, pp. 313–314). A common metaphor for motivated-blame theories is that people act like prosecutors whose ultimate goal is to mete out punishment rather than to discover the truth (Tetlock et al., 2007).

Evidence in support of the motivated-blame perspective suggests that, in the presence of initial negative moral evaluation, people are inclined to judge violations as intentional (e.g., Knobe, 2003), to inflate perceptions of harm (Ames & Fiske, 2013), to see perpetrators as more strongly causally involved (Alicke, 1992), and to exaggerate judgments of foreseeability (Mazzocco et al., 2004). This bias of seeing more intentionality, causality, harm, or foreseeability amounts to a tendency to embrace information that exacerbates blame and to discount information that mitigates blame. More explicitly, Alicke et al. (2011) write: "… the culpable-control model assumes that the control elements (behavior, causal, and outcome) that observers analyze are processed in a "blame validation" mode. Blame validation entails either exaggerating a person's actual or potential control over an event to justify the desired blame judgment or altering the threshold for how much control is required for blame" (p. 675).

Thus, the motivated-blame perspective predicts that the change of blame from the earliest possible judgment to the final assessment (after additional information has been processed), should be asymmetrically biased—favoring small reductions of blame in response to mitigating information (because it frustrates a desire to blame) but large surges of blame in response to exacerbating information (because it fulfills a desire to blame).

Predictions and Experiments

From the socially-regulated blame perspective, the burden of social warrant puts pressure on moral perceivers to have access to criterial information content (intentionality, reasons, and preventability), and the recently proposed Path Model of Blame describes in detail these information sources and their hierarchical relationships (Malle et al., 2014; Monroe & Malle, 2017). Applied to the case of judgment updating, the Path Model clarifies how an initial judgment of an ambiguous norm violation will be refined as more information becomes

available, making two sets of novel, theoretically-grounded predictions.[3] The first set of predictions concerns the *gradedness* of updates as a function of specific information sources (e.g., intentionality, justified reasons); the second set of predictions concerns the *symmetry* of mitigating vs. exacerbating updates.

Graded updating. A unique feature of blame according to the Path Model is its hierarchical organization of information processing. Applied to blame updating, the model predicts a two-step updating process (See Figure 1):

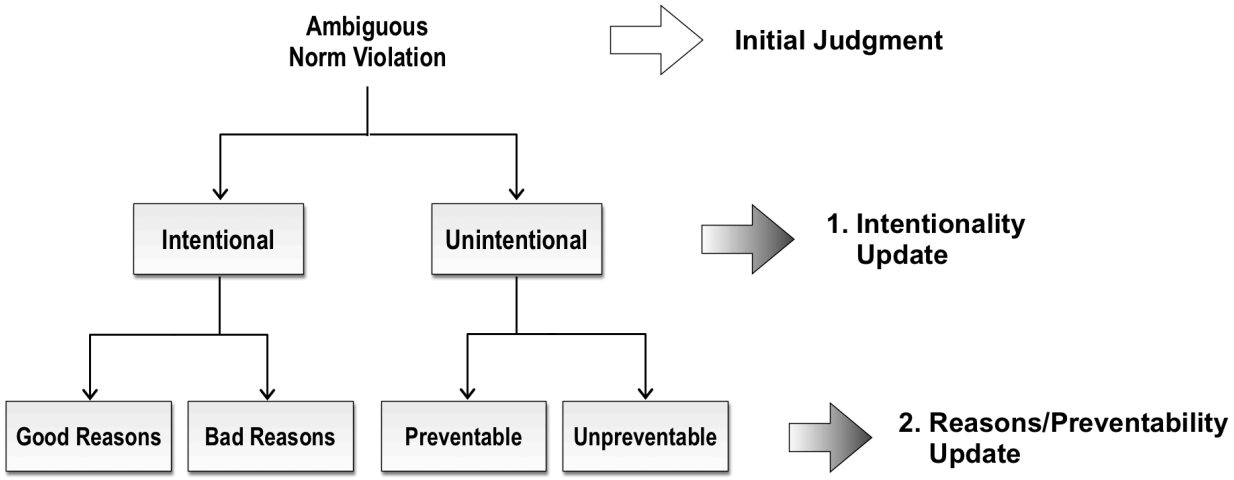

Figure 1. Two layers of blame updating after an initial judgment, according to the Path Model of Blame: The first update occurs after learning about the intentionality of the norm violation; the second update occurs after learning about the agent's reasons for the intentional violation or the preventability of the unintentional violation.

Blame judgments will be updated at a first level as it becomes clear whether the agent committed the violation intentionally or unintentionally. Similar to other models, the Path Model asserts that intentionality amplifies blame, which is already well supported in the literature (Darley & Shultz, 1990; Lagnado & Channon, 2008; Ohtsubo, 2007; Young & Saxe, 2009). The Path Model, however, makes the novel claim that intentionality judgments bifurcate moral information processing into two distinct tracks. On the intentional track, perceivers consider an agents' reasons for committing the violation; on the unintentional track, perceivers consider the preventability of the violation (Monroe & Malle, 2017). This is where the second level of updating occurs.

At this second level, blame judgments will be updated as it becomes clear either (a) whether the agent, if committing the violation intentionally, had good reasons or bad reasons for doing so, or (b) whether the violation, if committed unintentionally, was preventable or unpreventable for the agent. Because the second level provides additional information over the first, blame judgments can be updated in a graded manner. For example, relative to initial blame,

[3] The predictions developed here are not cast in terms of intuitive vs. deliberative processes. The Path Model explicitly makes room for both of these modes of processing (Malle et al., 2014, pp. 152, 156, 160, 177), and our methodology does not aim to differentiate between processing modes.

updated blame will increase when the violation proves to be intentional, but it will increase even more when that intentional violation was committed for bad reasons and decrease substantially when committed for good reasons. Similarly, updated blame will decrease when the violation proves to be unintentional, but it will decrease even more when that unintentional violation was unpreventable and decrease less so when the unintentional violation was clearly preventable.

More precisely, the Path Model of Blame makes three pairs of predictions regarding the gradedness of people's updated moral judgments:

(1) *Intentionality predictions:* Relative to initial blame for a violation whose intentionality is ambiguous, people will (a) decrease (mitigate) blame (↓↓)[4] when they learn that the violation was unintentional and (b) increase (exacerbate) blame (↑↑) when they learn that it was intentional.

(2) *Reasons predictions:* Beyond changes after learning only that a violation was intentional (↑↑), when people also learn the agent's specific reasons for the intentional violation, they will (a) increase blame further than for intentional-only if the agent had bad (unjustified) reasons (↑↑↑) but (b) substantially decrease blame compared with intentional-only if the agent had good (justified) reasons (↓↓↓).[5]

(3) *Preventability predictions:* Beyond changes after learning only that a violation was unintentional (↓↓), when people also learn about the unintentional violation's preventability, they will (a) decrease blame further than for unintentional-only if the agent could not have prevented the event (↓↓↓) but (b) decrease blame less than for unintentional-only if the agent could have prevented the event (↓).

Symmetric updating. The socially-regulated blame perspective and the Path Model of Blame lead to the hypothesis that, because blaming is an observable, costly, and socially-regulated act, perceivers should flexibly revise their blame judgments in response to new relevant evidence, whether that evidence supports increasing or decreasing blame. Inflexible updating would incur social costs—to the offender (when being blamed unfairly), the blamer (when found to have blamed unfairly), or third parties (e.g., when an offender gets away unsanctioned). These potential costs, and the community's interest in minimizing them, puts pressure on moral perceivers to be even-handed in updating their blame judgments in response to new information. This sets up a pair of predictions about symmetry within the present studies, beyond the gradedness predictions: (4) The blame mitigation in response to learning that an agent *unintentionally* caused harm will, on average, be of equal magnitude (↓↓) as the blame exacerbation in response to learning that an agent *intentionally* caused harm (↑↑). (5) The blame mitigation in response to an agent's *morally good* (justified) reasons for acting will, on average, be of equal magnitude as the blame exacerbation in response to learning about an agent's *morally bad* (unjustified) reasons for acting.

---

[4] The magnitude of "two arrows" is a reference point that allows additional grades of change (one and three arrows in either direction) that our hypotheses specify.

[5] The prediction that morally good reasons strongly mitigate blame may appear counterintuitive at first. However, previous studies have demonstrated the capacity for morally justified reasons to powerfully mitigate blame. For example, Greene et al. (2009) showed that agents' reasons shape people's moral judgments in trolley cases. Describing the switch-thrower's reasons for sacrificing one workman as an attempt to save the lives of the other five workmen makes the act appear morally permissible. Similarly, research focusing on both everyday and legal contexts shows that citing beliefs of feeling threatened and acting in self-defense justifies many forms of (even serious) harm (Finkel, Maloney, Valbuena, & Groscup, 1995; Robinson & Darley, 1995).

These predictions contrast with motivated-blame models. Although these models may in principle allow for gradedness predictions (although no extant model specifies them), the question of symmetry arguably differentiates the two perspectives. In particular, the motivated-blame perspective predicts that blame updating should generally be asymmetric—because of "observers' proclivity to favor blame versus non-blame explanations for harmful events and to de-emphasize mitigating circumstances"; Alicke, 2000, p. 565).

The postulated desire to blame should produce large blame surges (↑↑↑) in response to exacerbating information and relatively smaller blame reductions in response to mitigating information (↓). To our knowledge, no current motivated-blame model makes differential predictions about the impact of particular types of mitigating or exacerbating information; we therefore represent the increases and decreases of updated blame as uniform within exacerbation and mitigation, respectively (see Table 1).

We tested these predictions in six studies using a novel experimental paradigm of moral updating, in which people first receive a sparse description of a norm violation, make an initial blame judgment, receive additional information (that varies in mitigating or exacerbating contents), and make an updated judgment. Study 1 examined moral updating using a student sample and text stimuli. Study 2 recruited a community sample and contrasted the updating condition to a full-information control to evaluate whether people anchor on early blame judgments and asymmetrically adjust in response to mitigating vs. exacerbating information. Study 3 further tested this anchoring possibility by comparing the updating condition to a full-information control and a "silent" first judgment control condition. Study 4 replicated our core findings using audio stimuli. Lastly, Study 5 tested whether the predictions of graded blame change and symmetric updating were robust under cognitive load, and Study 6 tested whether the process of making and revising moral judgments is moderated by the transgressor's group membership.

Table 1. Blame change patterns in response to distinct pieces of new information, as predicted by the socially-regulated blame and motivated-blame models

	Blame change predicted by:	
New Information	**Socially-regulated blame model**	**Motivated-blame models**
Intentional only	↑↑	↑↑↑
Intentional with Bad Reasons	↑↑↑	↑↑↑
Intentional with Good Reason	↓↓↓	↓
Unintentional only	↓↓	↓
Unintentional but Preventable	↓	↓
Unintentional and Unpreventable	↓↓↓	↓

Note. Magnitudes of blame change are indicated by arrows (↑ for increase; ↓ for decrease). The number of arrows indicates ordinal differences in magnitude.

Statistical power, generalizability, and sample representativeness

For all studies, we report all manipulations and dependent measures. Each study's sample size and stopping rules were determined prior to data collection. Our studies use a within-subject design (with six-fold stimulus replication per design cell), and an a priori power analysis using G-power recommended a minimum sample size of 36 participants to detect a moderate effect size (partial $\eta 2 = .09$) with .9 power. Thus, across all of our studies we aimed to collect data from a minimum of 36 participants per condition. Further, we addressed power and the replicability of our findings in two additional ways. First, in Study 4 we substantially expanded sample size (n = 184) to increase power to 1.0. Second, to capture variation of effect sizes across experiments we conducted a meta-analysis of our core findings.

To examine the population generality of our findings we recruited three different samples across our studies. Studies 1 and 5 used student samples, Study 1 from a highly selective private university and Study 5 form a less selective public university. Participants in Studies 3, 4, and 6 were drawn from an internet sample using Amazon Mechanical Turk, where participants tend to be older, more diverse, and less educated than college student samples (Paolacci & Chandler, 2014). Finally, Study 2 used an adult community sample, which tended to be older compared to our college sample and had a level of education attainment that was representative of the United States (U.S. Census, 2015). Our findings replicate closely across these different participant samples and can be interpreted as generalizing broadly within the context of culturally Western populations.

Study 1

Method

Participants

Participants ($n = 60$) were students recruited from Brown University's subject pool. Two participants were omitted from the analyses for failing to complete the experiment (final $n = 58$). The sample was predominantly female ($n = 42$), and the majority of participants identified as White (57%), with fewer participants identifying as Asian (26%), Black (5%), Latin/Hispanic (2%), or multi-ethnic (7%). The sample had an average age of 19.5 years ($SD = 1.27$).

Procedure

Participants were tested in groups of two to six people. After participants provided informed consent, they were guided to individual testing rooms equipped with a desktop computer. The experimenter explained that the task involved reading brief descriptions of behavior on the computer and making judgments using an on-screen click-and-drag slider bar. Once participants indicated that they understood the task, the experimenter left the room and participants proceeded through a set of on-screen instructions and completed three practice trials. Then they completed 36 experimental trials divided into three blocks of 12, with a short break between blocks. After finishing the experimental trials, participants completed a brief demographics questionnaire and were debriefed.

Materials

Computer task. Each experimental trial consisted of four screens displayed in succession. Participants read a short description of a norm-violating event (screen 1, displayed for three seconds) and made an initial moral judgment ("How much blame does [agent] deserve?") using a click-and-drag slider bar with endpoints of 0 ("no blame at all") and 100 ("the most blame you would ever give") (screen 2). Immediately afterwards participants were presented with new information about the event along with the click-and-drag moral judgment slider bar and were free to update their initial judgment (screen 3). Finally, participants were asked to "write in their own words what happened" (screen 4) as a check of their understanding of the stimulus events. Participants were not allowed to revisit previous judgments or information.

Norm-violating event descriptions. Initial event descriptions were designed to cover a range of blameworthy behaviors from relatively minor harm (e.g., "Drew gave a customer incorrect change.") to severe harm (e.g., "Lisa shot Tom in the arm."). The event descriptions were designed to be ambiguous, containing only information about a moral agent, a patient, and a behavior—the minimal information components necessary for judgments of blame (Gray et al., 2012) (see Supplementary Materials for a list of behavior and pretest data).

Information updating. Following the initial moral judgment, participants were presented with one of six new pieces of information about the norm-violating event (see Supplementary Materials). This new information described whether the behavior was intentional or unintentional, whether the agent acted for morally good or bad reasons, or whether the agent could have foreseen and prevented the outcome or not. For example, for the initial event "Ted hit a man with his car," a participant would read one of the six types of new information described below:

1) **Intentional + morally bad reasons**: Ted intentionally hit a man with his car because he was in a hurry and did not feel like waiting on the man to cross the street.
2) **Intentional-only**: Ted intentionally hit a man with his car.
3) **Intentional + morally good reasons**: Ted intentionally hit a man with his car because he saw the man had a knife and was chasing a young, frightened woman.
4) **Unintentional + Preventable**: Ted accidentally hit a man with his car. Ted didn't check his blind spot before backing up.
5) **Unintentional-only**: Ted accidentally hit a man with his car.
6) **Unintentional + Unpreventable**: Ted accidentally hit a man with his car. Even though they were properly maintained, Ted's brakes failed to work.

The six types of new information were manipulated within-subjects, but any given participant saw only one new-information version of a given event narrative. In total, participants saw six replications of each type of new information, for a total of 36 events.

Updated blame judgments. To update their blame judgments after receiving new information, participants viewed the blame slider bar, with the pointer set at the position of the initial judgment, and had a chance to reposition it if so desired. To ensure that participants did not feel pressured to alter their initial judgments, instructions explicitly stated that they were not required to change their initial judgment. For each trial we recorded participants' updated blame judgments (i.e., the final position of the slider after participants confirmed their judgments) and then computed a change score of *updated blame – initial blame*.

Analysis. We tested the three pairs of gradedness predictions and the two symmetry predictions by defining the following within-subject contrasts. (1) *Intentionality predictions:* (a) Updated blame after people learn that the behavior was intentional (intentional-only trials) increases relative to initial blame; (b) updated blame after people learn that the behavior was unintentional (unintentional-only trials) decreases relative to initial blame. (2) *Reasons predictions:* (a) When people learn that the intentional behavior was performed for bad reasons blame further increases beyond intentional-only; (b) when people learn that the intentional behavior was performed for good reasons blame decreases relative to intentional-only. (3) *Preventability predictions:* (a) When people learn that the unintentional behavior was preventable blame decreases less than for unintentional-only; (b) when people learn that the unintentional behavior was unpreventable blame decreases more than for unintentional-only. (4) *Symmetry predictions:* (4) Blame updates (from initial to final) for intentional-only are indistinguishable in absolute magnitude from blame updates for unintentional-only. (5) Blame updates (from initial to final) for intentional actions performed for bad reasons are indistinguishable in absolute magnitude from blame updates for intentional actions performed for good reasons.

Results

The socially-regulated blame model predicts that blame change systematically decreases or increases as a function of an agent's intentionality, reasons, and preventability and that these changes are symmetric regardless of the information's mitigating or exacerbating content. A within-subject ANOVA revealed that new information content explained 84% of the variance in changed blame judgments, $F(5,285) = 305.0$, $p < .0001$, partial $\eta^2 = .84$, 95% CI [0.81, 0.86]. (See Figure 2). More specifically, each of the gradedness predictions was confirmed. (1) *Intentionality predictions:* Relative to initial blame for the ambiguous behavior, learning that the behavior was intentional exacerbated blame by 21.30 points, $t(57) = 18.70$, $p < .0001$, $d = 1.40$, 95% CI [0.86, 1.95]; learning that the behavior was unintentional mitigated blame by -19.22 points, $t(57) = -12.08$, $p < .0001$, $d = -1.10$, 95% CI [-1.54, -0.65]. (2) *Reasons predictions:* Learning the agent had morally bad reasons for acting further increased blame above intentionality alone ($M_{diff} = 3.72$), $t(57) = 2.98$, $p = .004$, $d = 0.44$, 95% CI [0.10, 0.77], but learning that the agent had morally good reasons substantially reduced blame compared to intentionality alone ($M_{diff} = -47.2$), $t(57) = 26.42$, $p < .0001$, $d = -4.41$, 95% CI [-6.09, -2.73]. (3) *Preventability predictions:* Learning that the violation was preventable reduced blame less than unintentionality alone ($M_{diff} = 5.78$), $t(57) = 3.53$, $p = .001$, $d = 0.53$, 95% CI [0.17, 0.88], but learning that it was unpreventable further reduced blame beyond unintentionality alone ($M_{diff} = -15.95$), $t(57) = 6.87$, $p < .0001$, $d = -1.07$, 95% CI [-1.58, -0.57].

Testing the symmetry predictions also showed support for the Path Model of Blame. (4) Comparing the absolute magnitude of blame change for the intentional-only and unintentional-only trials revealed that exacerbation in response to intentionality ($M = 21.3$, $SD = 8.67$) was symmetric with mitigation in response to unintentionality ($M = 19.2$, $SD = 12.1$), $t(57) = 1.18$, $p = .24$, $d = 0.20$, 95% CI [-0.22, 0.61]. (5) Likewise, comparing the absolute magnitude of blame change for morally good and bad reasons showed that exacerbation in response to morally bad reasons ($M = 25.0$, $SD = 8.34$) was symmetric with mitigation in response to morally good reasons ($M = 25.9$, $SD = 12.4$), $t(57) = -.45$, $p = .67$, $d = -0.09$, 95% CI [-0.44, 0.26].

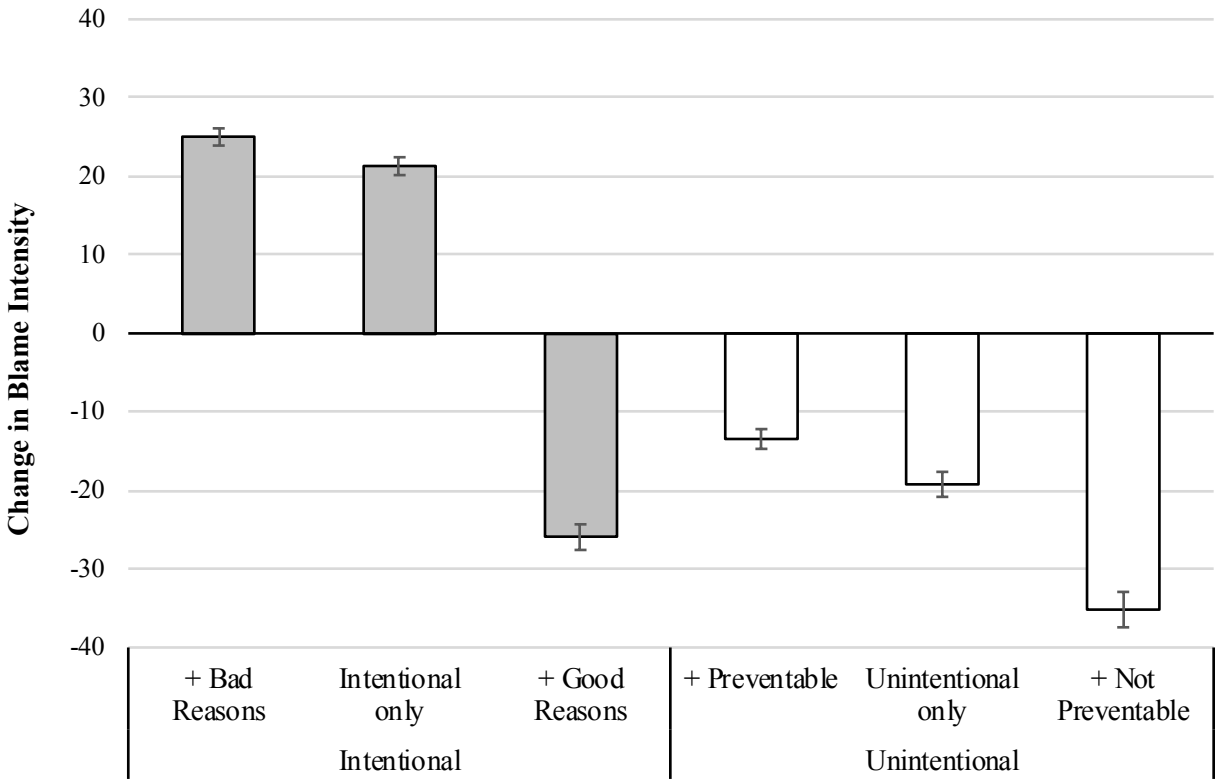

Figure 2. Blame change (relative to initial judgment) was a graded function of an agent's mental states, and mitigation (negative numbers) was symmetric with excerbation (positive numbers). Error bars = ±1 SE.

Discussion

The framework of socially-regulated blame suggests that, because blame evolved for social regulation and is subject to community norms, people are motivated to be relatively systematic in processing blame-relevant information. This systematicity should be particularly salient when people update their judgments, and the Path Model of Blame offers two sets of predictions of how people update blame in this circumstance. First, updates are predicted to be *graded* as a function of specific information sources (e.g., intentionality, justified reasons), and the results from Study 1 strongly support these predictions. Second, updates are predicted to be *symmetric* with respect to mitigating vs. exacerbating new information, and the results from Study 1 also strongly support these predictions. The latter finding stands in contrast to the motivated-blame perspective, which predicts diminished blame mitigation and enhanced blame exacerbation.

This study has three important limitations. First, it is unclear whether the evidential strength of good versus bad reasons and the convincingness of intentional versus unintentional behaviors were comparable. We therefore conducted a follow-up study (n = 120), which showed that, divorced from the context of making a blame judgment, people viewed information about

the morally bad reasons and about intentionality as actually more compelling.[6] This result makes any findings of symmetry particularly noteworthy. That is because, in isolation, the updated information, if anything, favored exacerbating blame (stronger bad reasons and more convincing intentionality) over mitigating blame.

A second limitation of Study 1 is that it relies on a sample drawn from a highly selective student population (see Henrich, Heine, & Norenzayan, 2010). Thus, it is possible that participants' seemingly systematic use of causal-mental information in moral updating reflects their capacity to reason more carefully than the general population. Lastly, one could argue that this study only partially tested the central claim of the motivated-blame models, which is that people anchor on their blame judgments and asymmetrically adjust these initial judgments when encountering new mitigating versus exacerbating information. Study 1 offered no criterion to evaluate the sufficiency of such adjustment.

In Study 2 we therefore recruited a community sample of participants, and we included a control condition in which participants make a single blame judgment with full information (combining the initial violation description and the subsequent information about intentionality, reasons, etc.), which thus precludes the impact of an anchor. Considering the level of blame in this full-information condition as the criterion, ordinary anchoring and insufficient adjustment in the two-part (initial-final) judgment condition would consist of overshooting updated blame if initial blame is high and undershooting updated blame if initial blame is low. Beyond that, motivated-blame processing would show *asymmetric* insufficient adjustment: overshooting when initial blame is high but not undershooting when initial blame is low—in the latter case, people should still substantially increase their blame judgments in order to satisfy the postulated desire for blame. The predictions derived from the Path Model remain the same as in Study 1: People should update their blame judgments as a result of the specific mitigating and exacerbating information they receive, whether information is presented at once (full-information condition) or in two parts.

Study 2

Study 2 addresses the limitations of Study 1 by using a community sample and by adding a full-information control condition to the design, which tests possible effects of anchoring and insufficient adjustment.

Method

Participants

Participants were recruited from the local RI community using a Craigslist ad for a "Paid research study." Participants who responded to the ad were invited into the lab to participate in the experiment. Participants were paid $20 for participating in the experiment. Out of 107 people who participated in the experiment, eleven were omitted from the analyses because they

[6] In the follow up study 60 participants rated how good or bad the agent's reason was (on an 11 point scale from - 5 extremely bad to 5 extremely good) and 60 participants rated how convinced they were that a given behavior was (as claimed) intentional/unintentional (on a -5 clearly unintentional to 5 clearly intentional scale). Paired *t* tests showed that people viewed bad reasons ($M = 3.41$ $SD = 1.36$) as significantly stronger than they viewed good reasons ($M = 1.45$ $SD = 1.04$), $t(59) = 10.81$, $p < .001$. Likewise, paired *t* tests demonstrated that participants were more convinced by intentional behaviors ($M = 3.51$ $SD = 1.55$) than by unintentional behaviors ($M = 1.54$ $SD = 1.84$), $t(59) = 9.25$, $p < .001$.

reported being unable to read. Participants were recruited separately, though simultaneously[7], for the updating (n = 58) and the full-information control condition (n = 38). Our target sample size for the updating condition was 60 participants (to match Study 1) and 40 participants for the full-information control condition.

Of the 96 participants who completed the experiment, the majority (n = 51) were male; 65% identified as White, 15% as Black, 8% as multi-ethnic, 6% as Latin/Hispanic, and 2% as Asian. Compared to Study 1, participants in this study were older (M = 32.5 years, SD = 12.4) and represented a diverse range of education. Forty-nine percent reported having a high school education only; 14.6% attained a 2-year degree; 24% attained a 4-year degree; and 11.4% attained a Master's degree or higher.

Procedure and Materials

Updating condition. Fifty-eight subjects comprised the updating condition. The stimuli, dependent variables, and procedure were identical to those in Study 1.

Full-information control condition. Thirty-eight subjects comprised the full-information control condition. The stimuli and dependent variables were identical to those in the updating condition, but in this condition participants received both the description of the norm-violating event and the mental or causal information in one sentence (e.g., "Tommy intentionally left the restaurant without leaving the waiter a tip because he didn't want to waste money on being nice."). After participants read this information they were asked to make a single blame judgment using the click-and-drag slider bar that ranges from 0 (no blame at all) to 100 (the most blame you would ever give).

Results

Updating Blame Judgments

A within-subjects ANOVA first tested the effects of new information on blame change within the updating condition, as a replication of Study 1. Overall, new information explained 70% of the variance of changes in blame judgment, $F(5, 285) = 133.5$, $p < .001$, partial $\eta^2 = .70$, 95% CI [0.64, 0.74] (see Figure 3). The specific gradedness predictions were largely confirmed. (1) Relative to the initial, ambiguous information, new information indicating that an event was intentional increased blame by 22.99 points (SD = 16.44), $t(57) = 10.65$, $p < .0001$, d = 1.29, 95% CI [0.75, 1.84]; new information indicating that an event was unintentional reduced blame by 19.51 points (SD = 21.25), $t(57) = -6.99$, $p < .0001$, d = -0.90, 95% CI [-1.32, -0.48]. (2) Unlike in Study 1, learning that an agent acted for morally bad reasons increased blame by only 0.94 points relative to learning that the agent acted intentionally, $t(57) = 0.43$, $p = .67$, d = 0.06, 95% CI [-0.21, 0.32]; however, replicating the pattern from Study 1, learning that an agent acted for morally good reasons reduced blame by 48.0 points compared to intentionality alone, $t(57) = 19.35$, $p < .0001$, d = -2.80, 95% CI [-3.89, -1.72]. (3) Preventable violations reduced blame by 6.85 points less than unintentionality alone, $t(57) = 2.21$, $p = .031$, d = 0.34, 95% CI [0.01, 0.67];

[7] Piloting revealed that the length of the experiment differed significantly between the updating condition (45 minutes) and the control condition (25 minutes). Because of the time difference between the two conditions, the IRB instructed that we advertise and recruit separately for the two conditions to ensure equitable compensation for participants.

unpreventable violations reduced blame by 14.29 points more than unintentionality alone, $t(57) = 4.22$, $p < .0001$, $d = -0.65$, 95% CI [-1.05, -0.26].

Additionally, supporting the symmetry predictions, corresponding pairs of exacerbating and mitigating information showed the same absolute magnitude of blame change. (4) Blame change for intentional ($M = 22.99$, $SD = 16.4$) and unintentional events ($M = 19.51$, $SD = 21.2$) were statistically indistinguishable, $t(57) = 0.96$, $p = .34$, $d = 0.18$, 95% CI [-0.18, 0.55]. (5) Likewise, blame change for morally good reasons ($M = 25.0$, $SD = 17.6$) and morally bad reasons ($M = 23.94$, $SD = 16.0$) were statistically indistinguishable, $t(57) = -0.31$, $p = .76$, $d = -0.06$, 95% CI [-0.39, 0.27].

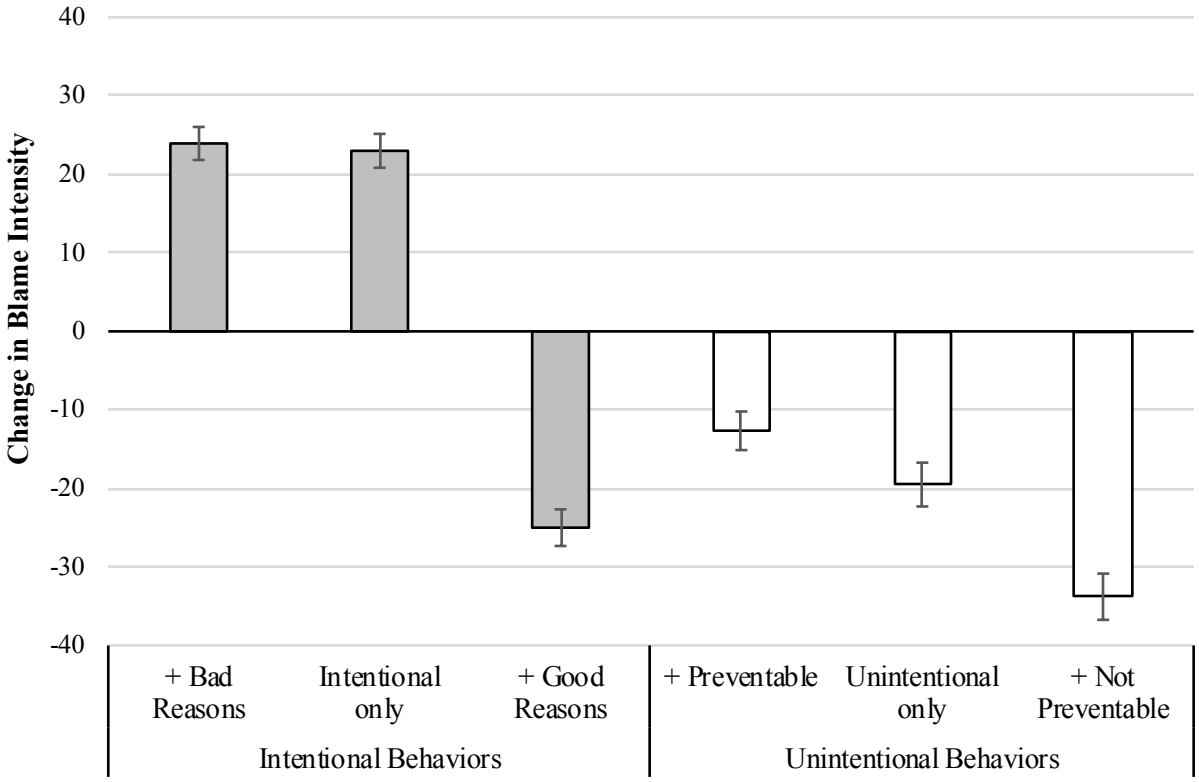

Figure 3. Blame change tracked agents' mental states for both mitigating an exacerbating information. Positive numbers indicate increases in blame from the initial judgment; negative numbers indicate decreases in blame judgments. Error bars = ±1 SE.

Comparing the Updating and Full-information Conditions

This study more directly contrasts predictions of the socially-regulated blame and the motivated-blame models by comparing the updating condition with a full-information control condition. A 2 (presentation condition) x 6 (new information) mixed between/within ANOVA showed that across both conditions, new information explained 72% of the variance in updated blame judgments, $F(5, 470) = 245.6$, $p < .001$, $\eta^2 = .72$, 95% CI [0.68, 0.75]. The motivated-blame perspective predicts that people in the updating condition, relative to the full-information condition, should show asymmetric anchoring patterns, leading to overall more blame for both exacerbation and mitigation. Contradicting this prediction, presentation condition had no impact

on average levels of blame, $F(1,94) = 1.46$, $p = .23$, $\eta^2 = .015$, 95% CI [0.00, 0.09]. There was a small information by condition interaction, suggesting that some changes due to new information were different in the two presentation conditions, $F(5, 470) = 3.24$, $p = .007$, $\eta^2 = .033$, 95% CI [0.003, 0.06]. We decomposed this interaction term into the hypothesis-relevant gradedness contrasts, asking whether any of these predictions was moderated by information presentation.

(1) The intentionality predictions were affected by presentation in a way that partially supported the motivated-blame prediction. The increase from initial to updated blame for intentional events was significantly higher in the updating condition ($M = 22.99$, $SD = 15.14$) than in the full-information condition[8] ($M = 12.08$, $SD = 18.88$), $t(94) = 3.00$, $p = .003$, $d = 0.65$, 95% CI [0.22, 1.07]. On the side of unintentional events, the results were inconclusive. The drop from initial to updated blame for unintentional events was weaker, but not significantly so, in the updating condition ($M = -19.51$, $SD = 22.96$) than in the full-information control condition ($M = -25.46$, $SD = 24.66$), $t(94) = 1.26$, $p = .21$, $d = 0.25$, 95% CI [-0.16, 0.67].

(2) The reasons predictions were also affected, but opposite to what the motivated-blame perspective would predict. The blame boost when learning that an agent acted for bad reasons (relative to learning merely that the agent acted intentionally) was smaller in the updating condition ($M = 1.90$, $SD = 14.13$) than in the full-information condition ($M = 10.08$, $SD = 16.72$), $t(94) = 2.58$, $p = .011$, $d = 0.54$, 95% CI [0.12, 0.96]. This suggests that when people update reason information they exacerbate blame relatively less than under full information. On the side of good reasons, presentation also had an impact, but again counter to the motivated blame perspective. The blame drop when learning that an agent acted for good reasons (relative to learning merely that the agent acted intentionally) was larger in the updating condition ($M = -55.96$, $SD = 22.62$) than in the full-information condition ($M = -38.67$, $SD = 19.99$), $t(94) = 3.81$, $p < .0001$, $d = 0.80$, 95% CI [0.37, 1.22]. This suggests that when people update reason information people mitigate blame *more* than under full information.

(3) The preventability predictions were not affected by presentation condition. The weaker blame mitigation when learning that an unintentional event was preventable (rather than merely unintentional) did not differ between the updating condition ($M = 7.84$, $SD = 21.90$) and the full-information condition ($M = 10.95$, $SD = 18.97$), $t(94) = 0.72$, $p = .48$, $d = 0.15$, 95% CI [-0.26, 0.56]. Similarly, the stronger blame mitigation when learning that an unintentional event was unpreventable (rather than just unintentional) did not differ between the updating condition ($M = -10.40$, $SD = 24.14$) and the full-information control condition ($M = -8.31$, $SD = 16.54$), $t(94) = 0.47$, $p = .64$, $d = 0.10$, 95% CI [-0.32, 0.51].

In sum, updated blame judgments were on average indistinguishable from blame under full information, but some variations across information types emerged. In two comparisons (one significant), updated blame patterns were consistent with what a motivated-blame hypothesis would suggest; in four comparisons (two significant) updated blame patterns were inconsistent with what a motivated-blame hypothesis would suggest.

---

[8] The full information condition had no "initial" blame ratings, so for the present analyses we rescaled updated blame judgments in both conditions by subtracting the specific initial blame ratings for each information condition (e.g., for intentional-only in the full-information group, subtract the intentional-only initial-blame rating from the updating group). This step makes it easier to interpret the results in terms of the sets of gradedness predictions, but statistically it corresponds to an analysis of the updated blame judgments as they are shown in Figure 4.

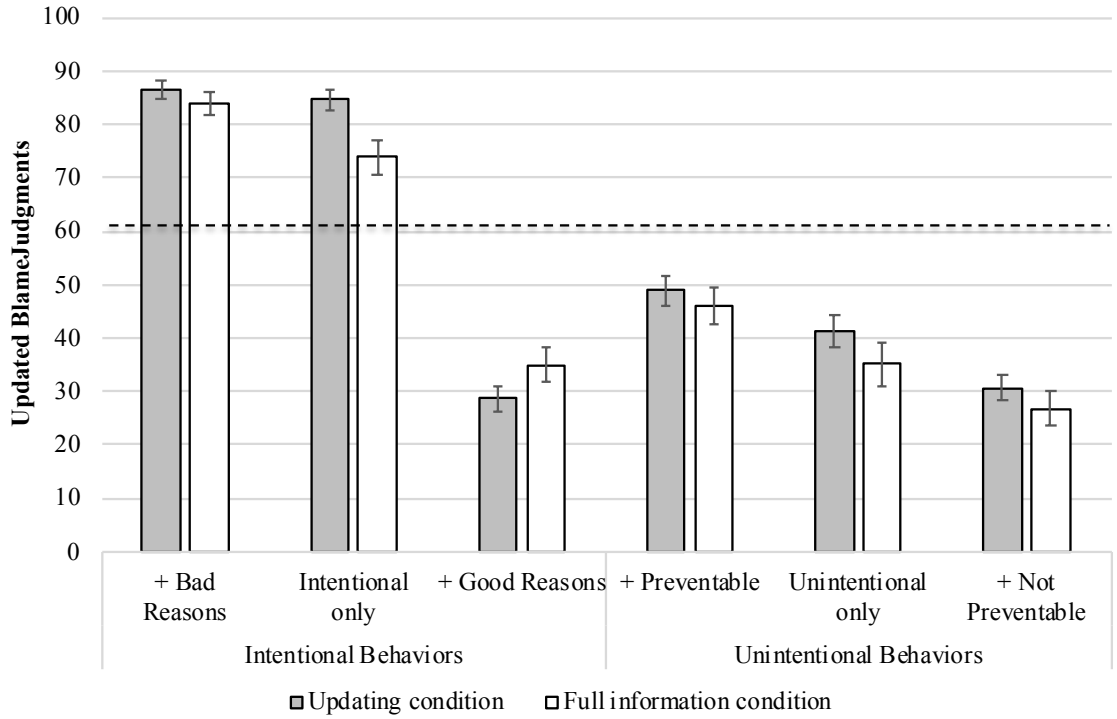

Figure 4. Final, updated blame judgments for the updating and the full-information control conditions. For comparison, dotted line indicates average initial blame in the updating condition ($M = 60.78$). Error bars = ±1 SE.

Discussion

This study closely replicated the results of Study 1. Using a community sample, we showed that people update blame judgments in a graded and symmetric fashion when new mental or causal information becomes available. Further, examining the pattern of blame change revealed the powerful impact of causal-mental state information on moral judgment consistent with previous research (Cushman, 2008; Lagnado & Channon, 2008; Martin & Cushman, 2016; Monroe & Reeder, 2011; Plaks, McNichols, & Fortune, 2009; Reeder, Monroe, & Pryor, 2008; Woolfolk, Doris, & Darley, 2006; Young & Saxe, 2009). By contrast, we did not find consistent evidence that people anchored and asymmetrically updated their moral judgments in favor of increasing or maintaining blame.

By comparing the updating condition with a full-information control condition we provided a more appropriate contrast of the two theoretical models and found further evidence for the socially-regulated blame perspective and the Path Model of Blame. Whether people made a single judgment or updated an initial judgment, they were comparably sensitive to mental state information and arrived at largely identical blame judgments. These data are suggestive of a flexible moral system that makes considerable adjustments in moral judgments if the available evidence favors such changes (see Mende-Siedlecki & Todorov, 2016 for similar findings in a non-moral domain). One exception to this broad pattern of findings was the comparison between updating in the "intentional" and the "intentional-with-bad-reasons" information conditions. Participants appeared to equate these two information conditions, which may reflect default

assumptions about intentional behavior being normally performed for bad reasons. That is, if one hears that "Steve intentionally punched Mark," one might suspect that Steve had morally unjustified reasons for this action. Whether such a default assumption constitutes a motivated bias is unclear, as we do not know the actual base rates of intentional norm violations performed for justified versus unjustified reasons.

One aspect of the present paradigm may have favored flexible, systematic updating: asking people to make explicit *initial* judgments. For one thing, demanding such a judgment might reduce the natural ambiguity of violations (because one has to commit to a certain construal); for another, demanding two public judgments (an initial and an updated one) might put pressure on people to show that they are properly taking the initial information into account. By itself, such experimental demand cannot explain the specific ordinal patterns of mitigated and exacerbated blame we found in Studies 1 and 2; but removing it would strengthen the interpretation that people spontaneously update their blame judgments in the predicted manner. Thus, Study 3 employed a modified updating condition in which people are invited to make a silent, undisclosed judgment in response to the ambiguous initial information and then make a single public judgment after the new information is revealed. Such a situation might also more closely resemble the everyday process of moral judgment, wherein people likely have initial moral reactions but do not express them until more information is presented.

Study 3

In Study 3 we compare the standard updating condition to a "silent" initial judgment condition and also employ a full information control condition, as in Study 2. According to the socially-regulated blame perspective, people's moral judgments are based on agents' mental state information regardless of whether early judgments are silent or explicit. The social demand for warrant (fair judgment based on evidence) is strong enough to guide people's initial judgment, the information updating process, and the public judgment. Thus, the socially-regulated blame model predicts that blame should change as a function of the agent's mental states, and updated blame judgments should show the same pattern of blame shown in Studies 1 and 2 across the updating, silent initial judgment, and the full-information conditions. Contrastingly, according to the motivated-blame perspective, a silent, unchecked initial judgment of a highly ambiguous scenario should activate people's desire to blame and drive a motivated-blame process when integrating new information, wherein they are more responsive to exacerbating information than mitigating information.

Method

Participants

Subjects (n = 120) were recruited from Amazon Mechanical Turk and randomly assigned to either an updating (n = 40), a full information (n = 40), or a silent initial judgment condition (n = 40). Average age in the sample was 35.2 years (SD = 10.9). The sample was evenly split between men (n = 59) and women (n = 58), with three participants declining to indicate their sex. The majority of participants identified as White (n = 94), with fewer people identifying as Black (n = 6), Latin/Hispanic (n = 7), Asian (n = 7), or multi-ethnic (n = 2).

Procedure and Materials

Procedures and materials were identical to Study 2 with one change: this study included a silent initial judgment condition in addition to the updating, and the full-information control condition (described above). In the silent initial judgment condition, each experimental trial consisted of three screens displayed in succession. In each trial, participants read a short description of a norm-violating event (screen 1) and were asked to "make a judgment (just in your own head) about the person. Once you've made this private judgment, click the button on the screen to move on." On the following screen (screen 2), participants read the updating information and were asked, "How much blame does [agent] deserve?" Participants made their responses using a click-and-drag slider bar with endpoints of 0 (no blame at all) and 100 (the most blame you would ever give). Finally, participants were asked to "write in a few words what happened" (screen 3) as a check of their understanding of the stimuli. As in previous studies, participants were not allowed to revisit previous judgments or information.

Results

A 3 (presentation condition) x 6 (new information) mixed between/within ANOVA showed that participants in all three information presentation conditions arrived at markedly similar blame judgments (Figure 5), and overall, new information explained 78% of the variance in updated blame judgments, $F(5, 585) = 414.6$, $p < .0001$, $\eta^2 = .78$, 95% CI [0.75, 0.80]. Presentation condition did not significantly affect average levels of blame, $F(2, 117) = 0.14$, $p = .87$, $\eta^2 = .002$, 95% CI [0.00, 0.03], and the presentation by new information interaction was marginally significant $F(10, 585) = 1.71$, $p = .076$, $\eta^2 = .028$, 95% CI [0.00, 0.04]. As in Study 2, we decomposed this interaction term into the hypothesis-relevant gradedness contrasts, asking whether any of these predictions was moderated by information presentation.

(1) Both of the intentionality predictions were confirmed, one with a moderation of information presentation. (a) Across information presentation conditions, participants' blame levels were higher in response to learning that an agent acted intentionally ($M_{diff} = 19.20$, $SD = 13.80$) than initial blame[9], $F(1,117) = 241.8$, $p < .001$, $d = 1.30$, 95% CI [1.11, 1.67]. This increase varied somewhat by information presentation, $F(2,117) = 3.46$, $p = .035$. Participants in the silent first judgment condition increased their blame judgments more strongly than participants in the full information condition ($p = .026$); however, participants in the updating condition did not significantly differ from either information presentation condition (ps > .38). (b) Examining unintentional events showed that, across information conditions, blame was reduced significantly (relative to initial blame) when participants learned that an agent behaved unintentionally ($M_{diff} = -27.43$, $SD = 21.95$), $F(1,117) = 184.8$, $p < .001$, $d = -1.25$, 95% CI [-1.53, -0.97]. This reduction was not affected by information presentation, $F(2,117) = 0.21$, $p = .81$.

(2) Both reasons predictions were confirmed regardless of information presentation. (a) Overall, when learning that the agents acted for morally bad reasons participants blamed them more ($M = 88.44$, $SD = 11.93$) than when learning merely that they acted intentionally ($M = 85.39$, $SD = 13.80$), $F(1,117) = 12.66$, $p < .001$, $d = 0.24$, 95% CI [0.09, 0.38]. This effect was

[9] Because the silent first and full information conditions had no known initial blame scores we subtracted the updating condition's corresponding initial blame score from all three conditions and effectively tested the resulting scores against zero.

not moderated by information presentation, $F(2,117) = 0.08$, $p = .93$. (b) Similarly, when learning that the agents acted for morally good reasons participant blamed them less ($M = 31.88$, $SD = 21.82$) than when learning merely that they acted intentionally ($M = 85.39$, $SD = 13.80$), $F(1,117) = 676.1$ $p < .001$, $d = -2.93$, 95% CI [-3.71, -2.15]. This difference was not moderated by information presentation, $F(2,117) = 2.26$, $p = .11$.

(3) Lastly, the preventability predictions were largely supported. (a) Across information presentation conditions, participants mitigated blame marginally less for unintentional preventable events ($M = 42.06$, $SD = 18.67$) than for merely unintentional events ($M = 39.24$, $SD = 21.95$), $F(1,117) = 2.88$, $p = .09$, $d = 0.14$, 95% CI [-0.02, 0.30]. This effect was not moderated by information presentation, $F(2,117) = 0.27$, $p = .76$. (b) Moreover, people mitigated blame more for unpreventable unintentional events ($M = 22.42$, $SD = 18.17$) than for merely unintentional events ($M = 39.24$, $SD = 21.95$), $F(1,117) = 71.15$, $p < .0001$, $d = -0.83$, 95% CI [-1.12, -0.55]. This effect was also not moderated by information presentation, $F(2,117) = 0.41$, $p = .67$.

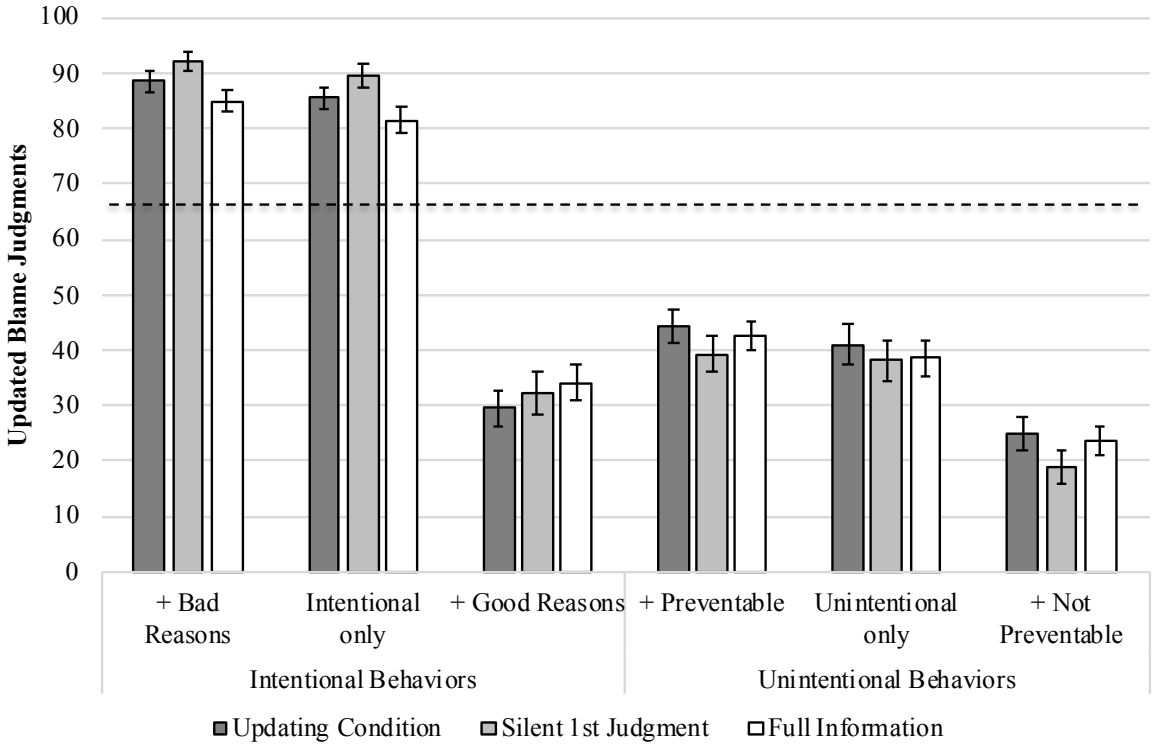

Figure 5. Updated blame judgments reliably tracked mental state information across the updating, silent first judgment, and full information conditions. Dotted line indicates average initial blame for updating condition ($M = 66.87$). Error bars = ±1 SE.

Discussion

Study 3 offered further support for the socially-regulated blame framework and the predictions of the Path Model of Blame. Mental and causal information systematically influenced moral judgments, and this effect held largely regardless of whether people updated an initial moral judgment, made a single judgment with full information, or made a private

judgment prior to learning about the agent's mind. Under the condition most conducive to motivated blame (silent first judgment), there was no evidence of muted mitigation, only one case of slightly stronger blame when the behavior was intentional.

Together, the first three studies demonstrate that people reliably update blame judgments in a graded and symmetric fashion. That is, people systematically update their judgments of blame as a function of an agent's mental states (e.g., intentionality, justified reasons), even making such fine-grained distinctions as between intentionally harming compared to intentionally harming for morally bad versus good reasons. Further, comparing moral updating to a full-information control condition, we find no evidence that people's initial moral judgments provide anchors that bias them towards asymmetric adjustments (such as more exacerbation or less mitigation).

In Study 4 we sought to demonstrate the robustness of these findings under tighter stimulus exposure conditions by converting all of the experimental stimuli (i.e., initial event descriptions and new information) into audio stimuli. The use of audio stimuli removes nonsystematic variation due to reading times and allays concerns that participants may preferentially revisit mitigating information in the experimental context, allowing them to process such information more carefully than they would under normal circumstances. In addition, in this replication we recruited a sample three times larger than the updating condition samples in the first three studies.

Study 4

Method

Participants

Participants (n = 200) were recruited from Amazon Mechanical Turk. Sixteen participants failed to complete the experiment and were omitted from the analyses (final n = 184)[10]. The majority of participants were female (58%) and white (75%), with smaller numbers of participants identifying as Black (9%), Latin/Hispanic (9%), Asian (3%), or multi-ethnic (2%). The average age of participants was 32.9 years (SD = 10.3).

Procedure and Materials

Procedures and materials were identical to Study 1 with one exception. In the current study, the initial and new information were presented as audio streams (rather than as on-screen text). The audio stimuli were 2-4 seconds long, recorded with neutral affect by a female speaker voice who had no knowledge of the research hypotheses. After each audio segment, the program automatically advanced to the relevant judgment screen (either initial blame or updated blame). Thus, participants listened to a short description of a norm-violating event (Screen 1), made an initial moral judgment (Screen 2), listened to the updating information (Screen 3), and finally had an opportunity to update their blame judgment (screen 4). Participants were not able to return to previous screens.

[10] The 16 participants who were removed from the analyses completed fewer than 20% of the experimental trials. All data removal was conducted before data analysis.

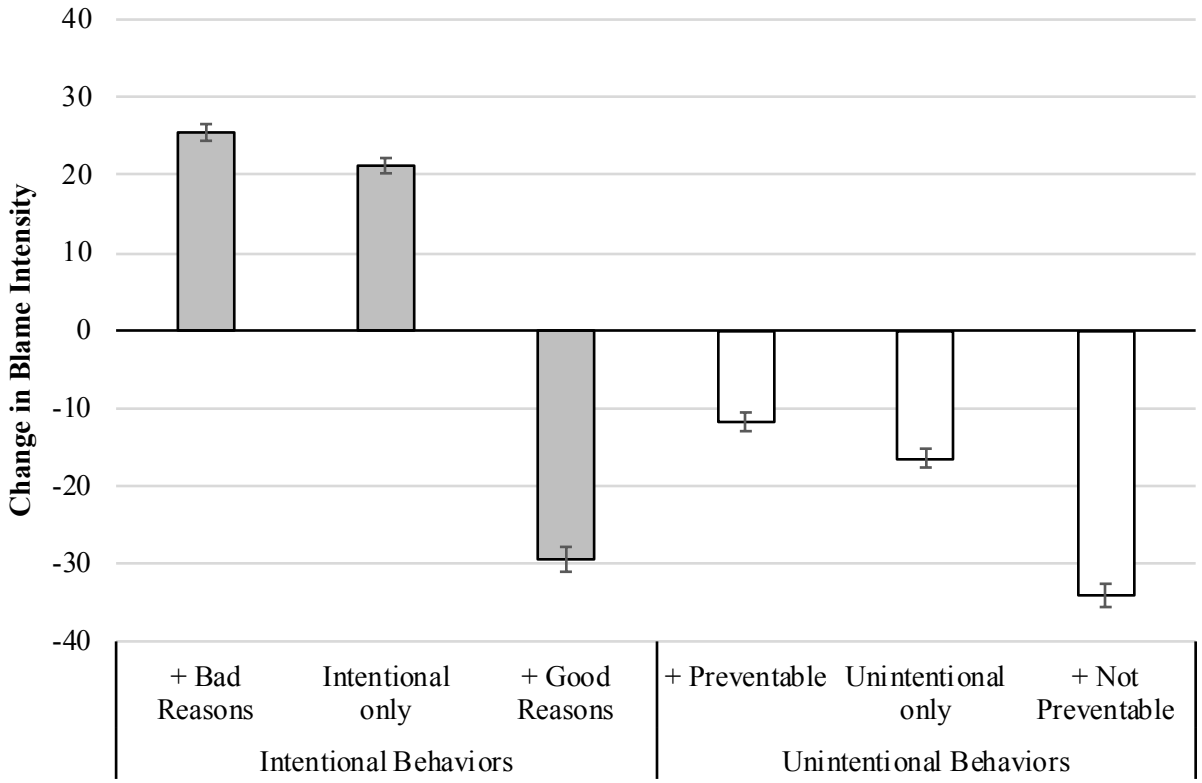

Figure 6. Large-sample replication with audio stimuli: Perceivers make fine-grained distinctions in updated blame as a function of new causal or mental information, in line with gradedness and symmetry predictions. Positive numbers indicate increases in blame from the initial judgment; negative numbers indicate decreases in blame judgments. Error bars = ±1 SE.

Results

Replicating previous studies, new information explained 75% of the variance in changed blame judgments, $F(5,915) = 548.8$, $p < .001$, partial $\eta^2 = .75$, 95% CI [0.72, 0.77] (See Figure 6). Also, all of the gradedness predictions were confirmed. (1) Relative to initial blame, learning that an event was intentional increased people's blame by 21.20 points, $t(183) = 21.58$, $p < .0001$, $d = 1.32$, 95% CI [1.03, 1.62]; learning that it was unintentional reduced their blame by 16.42 points, $t(183) = -13.55$, $p < .0001$, $d = -0.84$, 95% CI [-1.05, -0.63]. (2) Both reasons predictions were confirmed. Learning that an agent had morally bad reasons for acting increased people's blame by 4.27 points over learning merely that the agent acted intentionally, $t(183) = 4.58$, $p < .0001$, $d = 0.31$, 95% CI [0.16, 0.45]; learning that the agent had morally good reasons for acting reduced blame by 50.64 points, $t(183) = 30.68$, $p < .0001$, $d = -2.81$, 95% CI [-3.42, -2.21]. (3) Both preventability predictions were confirmed. Learning that a violation was preventable reduced blame by 4.67 points less than unintentionality alone, $t(183) = 4.08$, $p < .0001$, $d = 0.28$, 95% CI [0.14, 0.43]; learning that a violation was unpreventable reduced blame by 17.67 points more than unintentionality alone, $t(183) = 12.25$, $p < .0001$, $d = -0.96$, 95% CI [-1.21, -0.71].

We also examined the symmetry predictions using corresponding pairs of exacerbating and mitigating information. (4) In contrast to the previous studies, blame exacerbation for

intentional violations (M = 21.20, SD = 13.33) was larger than blame mitigation for unintentional events (M = 16.42, SD = 16.44), $t(183) = 4.36$, $p < .001$, $d = 0.32$, 95% CI [0.11, 0.53]. (5) Going in the opposite direction, however, blame mitigation for morally good reasons (M = 29.44, SD = 21.71) was larger than blame exacerbation for morally bad reasons (M = 25.47 SD = 14.57), $t(183) = -2.96$, $p = .003$, $d = -0.21$, 95% CI [-0.40, -.03]. Thus, while the first finding is in line with a motivated blame perspective, the second runs contrary to this perspective. As a result, the absolute magnitudes of change across all mitigating (M = 22.92, SD = 13.83) and exacerbating information trials (M = 23.34, SD = 12.45) were indistinguishable, $t(183) = 0.27$, $p = .787$, d = 0.03, 95% CI [-0.14, 0.21].

Discussion

Study 4 offers further evidence for the predictions of the Path Model of Blame and the broader framework of socially-regulated moral judgment. People's blame judgments reliably and systematically tracked the mental states and causal contributions of agents, differentiating in a graded and symmetric way between different kinds of mitigating and exacerbating information updates.

Table 2. Meta-analysis of three pairs of gradedness predictions and a pair of symmetry predictions across Studies 1 through 4

Predictions	Effect size (d)	95% CI of d	z	p	Q	p
Intentionality predictions						
Intentional (Updated vs. Initial)	**1.49**	1.25, 1.74	11.91	< .001	0.31	.96
Unintentional (Updated vs. Initial)	**-0.96**	-1.15, -0.77	-10.06	< .001	2.82	.42
Reasons predictions						
Bad Reasons (vs. Intentional-only)	**0.25**	0.10, 0.39	3.30	< .001	4.11	.25
Good Reasons (vs. Intentional-only)	**-3.06**	-3.65, -2.49	-10.36	< .001	3.59	.31
Preventability predictions						
Preventable (vs. Unintentional-only)	**0.29**	0.17, 0.41	4.83	< .001	3.05	.38
Unpreventable (vs. Unintentional-only)	**-0.88**	-1.06, -0.70	-9.60	< .001	2.39	.50
Symmetry Predictions (Updated vs. Initial)						
Intentional-Unintentional Symmetry	0.12	-0.16, 0.40	0.85	.390	7.72	.052
Bad Reasons-Good Reasons Symmetry	-0.25	-0.52, 0.01	-1.85	.064	7.84	.049

Note. All result are based on random effects models across four studies, with inverse variance weights to correct for imprecision. Q is the degree of heterogeneity among studies, tested against the null hypothesis of no heterogeneity. Boldfaced entries are ones that were predicted to differ from zero in the indicated directed.

Meta-Analysis

To demonstrate the consistency of findings from Studies 1 through 4 we conducted a meta-analysis of the effect sizes for the three pairs of gradedness predictions and the pair symmetry predictions. Table 2 shows that all six gradedness predictions were supported, with sufficient homogeneity across the four studies, despite varying participant populations and varying stimulus presentations. Likewise, we found evidence for the two symmetry hypotheses, in the sense that a motivated-blame prediction of greater exacerbation than mitigation could not be supported (the reasons symmetry test tended to go even in the opposite direction, suggesting greater mitigation than exacerbation).[11]

Study 5

Despite these consistent results, the natural context of everyday blaming may involve cognitive distractions or a lack of motivation to engage in effortful cognition, which the previous studies did not incorporate. Thus, the studies may have inflated the evidence for our predictions because they allowed participants ample time and resources to engage in the kind of effortful cognition that is necessary to override otherwise influential biases. We tested this possibility in Study 5 by reducing the cognitive resources available to participants and examining whether the predictions of graded blame change and symmetric updating were robust under cognitive load. The socially regulated blame perspective implies such robustness, because the repeated demand to have warrant for blame judgments makes the requisite information processing of causal and mental information fast and effortless (Barrett, Todd, Miller, & Blythe, 2005; Decety & Cacioppo, 2012; Gray & Wegner, 2008; Malle & Holbrook, 2012; Monroe & Malle, 2017).

We manipulated cognitive load by randomly assigning participants to update their moral judgments either as usual or while producing a series of random taps with their index finger. Previous research has shown the tapping task to be effective at targeting executive functioning, while leaving secondary mechanisms, such as the phonological loop, unaffected (Stuyven, Van der Goten, Vandierendonck, Claeys, & Crevits, 2000). Because the socially regulated blame perspective predicts no change in information processing as function of cognitive load, we included a Stroop task as a manipulation check for the cognitive load manipulation. In this way, we could ascertain that the tapping task successfully placed participants in a state of cognitive load (slower responses and more errors on the Stroop task) independent of whether cognitive load alters people's updated blame judgments.

Method

Participants

Participants (n = 80) were students recruited from Florida State University's psychology subject pool. Two participants failed to complete the experiment and were omitted from the analyses (final n = 78). Participants were randomly assigned to cognitive load condition ($n_{standard}$ = 38; n_{load} = 40).

[11] There is a seeming discrepancy between the fact that we found symmetry between intentional and unintentional updating but the individual (absolute) effect sizes appear to differ: d = 1.49 for intentional-only and 0.96 for unintentional-only. This may be explained by the fact that the intentional-only condition often had smaller standard deviations and largely similar mean differences therefore received better d values.

Procedure and Materials

Procedures and materials were identical to Study 4 except that the current study included a between-subjects manipulation of cognitive load. In the standard condition, participants listened to a description of a norm-violating event, made an initial moral judgment, heard new information about the event, were invited to update their moral judgment, and finally typed their description of the event. In the cognitive load condition, participants listened to the event description and made their initial moral judgments as usual; however, participants were instructed to produce a series of random taps with their left hand during the presentation of the new information and while updating their judgments. In addition to the magnitude of blame judgments we also recorded response times for those judgments to determine the possible impact of cognitive load.

The cognitive load manipulation. The cognitive load task, labeled the "tapping task," was adapted from Stuyven et al. (2000). Participants were asked to produce a series of random taps with their left index finger on the desk. They were instructed to tap at least once every second and to try to be as random as possible. With the experimenter present, participants listened to two audio examples of random taps and practiced tapping during two updating trials. After the practice trials, the experimenter gave every participant verbal feedback, saying "That was good, but try to be even more unpredictable during the experimental trials. Remember that you need to focus on making your tapping as random as possible." The experimenter then left the room and participants completed the moral judgment task. As in the standard condition, participants listened to the event description and made their initial moral judgments. Prior to the presentation of the new information, a male voice instructed them to "Begin tapping." Participants tapped with their left hands during the presentation of the new information and while they updated their blame judgments. After making the updated judgment, the same male voice instructed participants to "Stop tapping," and then participants were asked to type a brief description of the event.

The Stroop task. After the moral judgment task participants completed the Stroop task. They were asked to make speeded responses to on-screen stimuli. In each trial, participants were presented with a word (BLUE, RED, YELLOW, or GREEN) written in different colors of font (Blue, Red, Yellow, or Green). They were told that if the word appeared in blue font to press "B;" if the word appeared in a red font to press "R;" and similarly for words in green (press "G") and yellow (press "Y"). For congruent trials the word (e.g., GREEN) matched the font color, whereas for incongruent trials it did not (e.g., GREEN appeared in red font). Performance in the Stroop task was measured by response time (RT) and number of errors (e.g., pressing "G" for the word GREEN displayed in blue font) in the incongruent trials relative to the congruent trials.

Participants who had been assigned to the cognitive load condition for the moral judgment task also experienced cognitive load in the Stroop task, tapping randomly with their left hand while pressing letters on the keyboard to indicate the stimulus words' font color[12].

[12] Stroop data from five participants was lost due to an experimenter error. Thus, the dataset for the Stroop task includes 36 participants in the control condition, and 37 participants in the cognitive load condition.

Results and Discussion

Manipulation Check: The Stroop task

A 2 condition (cognitive load vs. standard) x 2 trial type (congruent vs. incongruent) mixed between/within ANOVA revealed the predicted main effect of cognitive load on participants' response times. Regardless of trial type, participants under cognitive load labeled the font colors more slowly (M = 1119 ms, SD = 189 ms) than participants in the standard condition (M = 1019 ms, SD = 184 ms), $F(1,71) = 5.22$, $p = .025$, $d = 0.54$, 95% CI [0.07, 1.00]. Further, an analysis of the participants' response errors showed that cognitive load caused participants to make nearly twice as many errors (M = 1.97, SD = 1.76) as participants in the standard condition (M = 1.08, SD = 1.30), $t(71) = 2.45$, $p = .016$, $d = 0.58$, 95% CI [0.11, 1.04]. In addition to these effects of cognitive load, there was also the familiar "Stroop effect" of slower RTs for incongruent (M = 1191 ms, SD = 246 ms) than congruent trials (M = 947 ms, SD = 195 ms), $F(1,71) = 83.9$, $p < .001$, $d = 1.08$, 95% CI [0.79, 1.36]. There was no condition by trial type interaction, $F(1,71) = 1.44$, $p = .23$, $d = 0.28$, 95% CI [-0.18, 0.74].

Testing the Effects of Cognitive Load on Moral Updating

A 2 (cognitive load) x 6 (new information) mixed between/within ANOVA revealed that new information explained 79% of the variance of blame change, $F(5,380) = 289.0$, $p < .001$, partial $\eta^2 = .79$, 95% CI [0.76, 0.82] (see Figure 7). By contrast, the cognitive load manipulation did not affect blame change, $F(1,76) = 1.17$, $p = .282$, partial $\eta^2 = .015$, 95% CI [0.00, 0.11], and there was no load by information interaction, $F(5,380) = .331$, $p = .894$, partial $\eta^2 = .004$, 95% CI [0.00, 0.01]. Thus, cognitive load did not alter the systematic ways people used new information in updating their moral judgments.

The cognitive load manipulation did, however, significantly affect the speed with which people updated their blame judgments, $t(76) = 2.22$, $p = .029$, $d = 0.50$, 95% CI [0.05, 0.95]. Participants under cognitive load took significantly longer (M = 4260ms, SD = 1165) than participants in the control condition (M = 3729ms, SD = 926). Thus, cognitive load did reduce the efficiency of moral judgment responses but not the systematic manner in which people adjusted their moral judgments in response to new information.

As cognitive load did not affect blame change and did not interact with updating information, we averaged across the cognitive load and the standard conditions when evaluating the gradedness and symmetry predictions.[13] Replicating our previous studies we confirmed all of the gradedness predictions. (1) For intentionality, the data show that, relative to initial blame, intentional events exacerbated blame by 20.55 points, $t(77) = 15.86$, $p < .0001$, $d = 1.29$, 95% CI [0.85, 1.73]; whereas unintentional events mitigated blame by 17.11 points, $t(77) = -10.90$, $p < .0001$, $d = -0.82$, 95% CI [-1.12, -0.53]. (2) Acting for morally bad reasons increased blame 7.23 points above intentionality alone, $t(77) = 5.00$, $p < .0001$, $d = 0.59$, 95% CI [0.29, 0.89], and acting for morally good reasons reduced blame by 49.67 points compared to intentionality alone, $t(77) = 23.10$, $p < .0001$, $d = -3.05$, 95% CI [-4.06, -2.04]. (3) Lastly, learning that a violation was preventable reduced blame by 6.52 points less than unintentionality alone, $t(77) = 2.21$, $p = .030$, $d = 0.38$, 95% CI [0.07, 0.69], and learning that a violation was unpreventable reduced blame by 20.54 points more than unintentionality alone $t(77) = 9.38$, $p < .0001$, $d = -1.32$, 95% CI [-1.82, -0.82].

[13] Effects broken out by cognitive load vs. control are indistinguishable from the overall analysis.

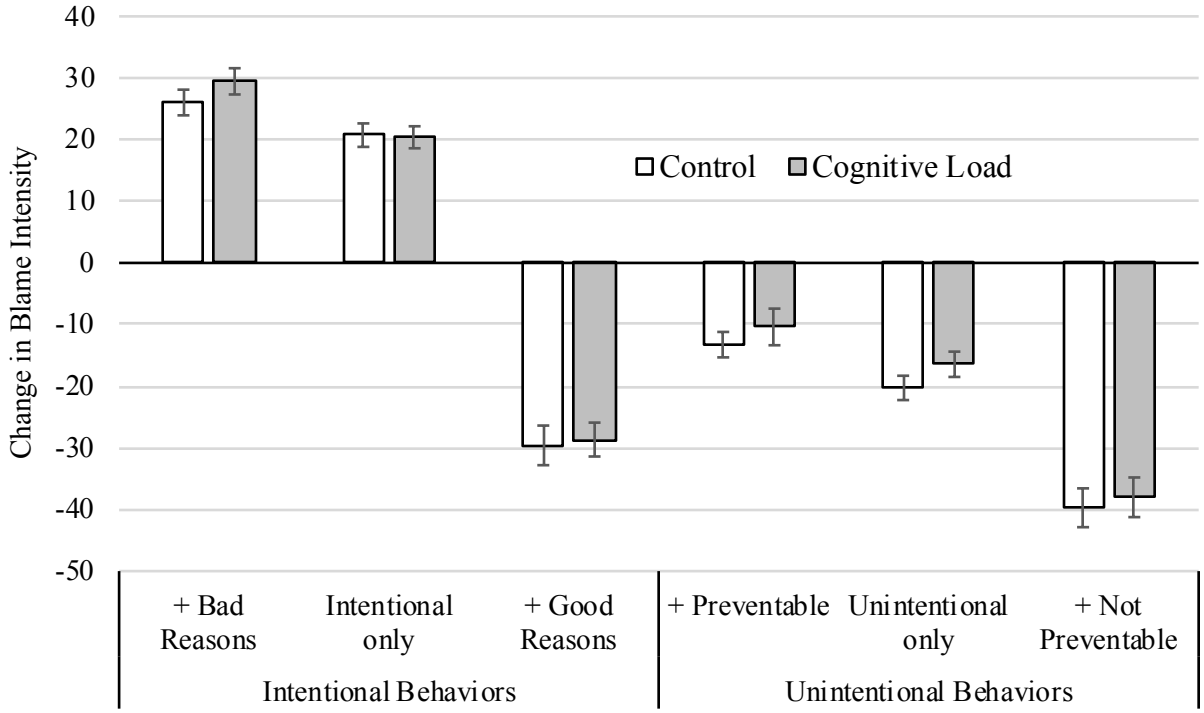

Figure 7. Patterns of blame change are guided by the systematic use of causal mental state information in both the control and cognitive load conditions. Error bars = ±1 SE.

Lastly we found partial support for the two symmetry predictions. (4) For intentionality, blame change for intentional events (M = 20.55, SD = 11.44) was slightly larger than change for unintentional events (M = 17.11, SD = 13.87), $t(77) = 1.77$, $p = .08$, $d = 0.28$, 95% CI [-0.07, 0.62]. (5) The magnitudes of blame change for morally good reasons (M = 29.12, SD = 18.49) and morally bad reasons (M = 27.79 SD = 13.20) were indistinguishable, $t(77) = -0.74$, $p = .36$, $d = 0.08$, 95% CI [-0.34, 0.18].

Discussion

Study 5 tested the hypothesis that flexible time and ample resources in the previous studies may have allowed participants to engage in graded and symmetric moral updating. Under more limited resources brought about by cognitive load, this systematic processing should suffer and potentially reveal deeper tendencies for motivated blame. Results revealed that although the load manipulation successfully limited people's cognitive resources (as evidenced by impaired Stroop task performance), causal and mental information remained a powerful and systematic determinant of blame change. This finding highlights the robustness of the socially-regulated blame framework and the predictions of the Path Model. Further, this study suggests that the findings in Studies 1-4 capture people's default method for arriving at (and revising) moral judgments rather than a process benefitting from an artificial or cognitively permissive paradigm.

Study 6

In multiple studies we have now demonstrated consistent evidence for the predictions of the Path Model of Blame and the broader framework of socially-regulated blame. Despite robust evidence of systematic moral updating, however, it is unlikely that the process of making and revising moral judgments is impervious to bias. In fact, the socially regulated blame perspective suggests one factor that may instill such bias: the norm transgressor's group membership. Demand for warrant in moral judgments is the pressure that a community puts on its members to maintain fair and justified moral regulation; however, such pressure may not extend to members of outgroups. Indeed, previous research has documented that people infer negative or sinister motives for the ambiguous behavior of political outgroups (Hulsizer, Munro, Fagerlin, & Taylor, 2004; Munro, Weih, & Tsai, 2010; Reeder, Pryor, Wohl, & Griswell, 2005), and outgroup members tend to be punished more harshly than ingroup members (Lieberman & Linke, 2007; Schiller, Baumgartner, & Knoch, 2014; Tajfel, 1970).

Therefore, in Study 6 we tested whether a transgressor's outgroup identity limits people's systematic moral updating and reveals motivated bias in causal and mental information processing. In particular, we examined whether the gradedness and symmetry of blame judgments are altered when judging outgroup transgressors. To manipulate ingroup/outgroup status we recruited strongly identified political partisans (Democrats or Republicans) and asked them to complete a modified moral updating experiment. After each ambiguous event description participants made their initial moral judgments, then learned about the transgressor's political affiliation (Democrat for half of the trials, Republican for the other half), and received, as usual, the new causal or mental state information. Thus, the transgressor's ingroup/outgroup status was a within-subject factor, tailored to the two participant groups of political supporters.

Method

Participants

We recruited 120 participants (60 self-identified Democrats and 60 Republicans) for a study entitled "Politics and Moral Judgment" using Amazon Mechanical Turk. Four participants failed to complete the experiment and were omitted from the analyses. To obtain our Democrat and Republican sample, we posted two identical HITs, one recruiting self-identified Democrats (n = 58) and the other recruiting self-identified Republicans (n = 58). The HITs were posted simultaneously, and participants could only complete one version of the study. Each HIT had a stopping rule of n = 60.

The samples were comparable in age (M_{Rep} = 38.3, SD = 13.6; M_{Dem} = 37.2, SD = 13.2), though other demographic differences emerged. There were fewer women in the Republican sample (n = 23) than in the Democratic sample (n = 30), and the Republican sample had a higher proportion of participants who identified as white (88%, n = 51) compared to the Democratic sample (67%, n = 39), though neither of these differences attained conventional statistical significance (p = .151 and p = .074, respectively). On a scale from 0 (Not at all religious) to 6 (Very religious), Republican participants were more religious (M = 4.00, SD = 2.03) than Democratic participants (M = 1.67, SD = 1.96), $t(114) = 6.28$, $p < .001$, d = 1.02. Republicans also identified marginally more with their party (M = 5.16, SD = 0.99) compared to Democrats (M = 4.72 SD = 1.42), $t(113) = 1.91$, p = .059, d = 0.36 (Scale: 0 = Not at all; 6 = Very strongly).

Procedure and Materials

Participants completed a moral updating task modified from Study 1. Participants first learned about an ambiguous immoral event (e.g., "Tommy left the restaurant without leaving the waiter a tip.") and made an initial moral judgment using a click-and-drag slider bar with endpoints of 0 (no blame at all) and 100 (the most blame you would ever give).[14] Immediately afterwards participants were presented with the updating information, which included the transgressor's political affiliation (e.g., "Tommy, a long-time Democrat, accidentally left the restaurant without leaving the waiter a tip."). Participants were then allowed to update their initial blame judgment. After participants submitted their blame judgment they were asked to recall the transgressor's political affiliation ("What political party does Tommy belong to?"). Participants completed a total of 12 updating trials (6 Democrat transgressors, 6 Republican transgressors).

Results

We conducted a 2 (transgressors' ingroup vs. outgroup membership) × 6 (new information) within-subjects ANOVA on blame change scores. New information explained 55% of variance in changed blame judgments, $F(5,575) = 139.7$, $p < .001$, $\eta^2 = .55$, 95% CI [0.49, 0.59]. Introducing the specter of motivated bias, transgressors' group membership also explained 7% of people's average blame change, $F(1,115) = 8.78$, $p = .004$, $\eta^2 = .071$, 95% CI [0.01, 0.17]. Specifically, when the transgressor was a political outgroup rather than ingroup member, participants' updated blame judgments were overall 5.23 points higher, $d = -0.31$, 95% CI [-0.53, -0.09]. By contrast, the group membership by information interaction was not significant, $F(5,575) = 0.70$, $p = .622$, $\eta^2 = .006$, 95% CI [0.000, 0.015]. Thus, it appears that group membership affects people's general intensity of blame, but not the graded adjustments in response to differential information. However, this omnibus test with five degrees of freedom may conceal effects in specific information conditions, so we tested the interaction term for each of the three pairs of predictions, following the overall prediction tests reported below.

Gradedness predictions. (1) Compared to the ambiguous initial event descriptions, events described as merely intentional increased blame by 24.56 points, $t(115) = 12.84$, $p < .0001$, $d = 1.19$, 95% CI [0.83, 1.56]. This pattern did not vary by group identity, $t(115) = 0.71$, $p = .48$. Similarly, events described as merely unintentional reduced blame by 19.99 points, $t(115) = -8.89$, $p < .0001$, $d = -0.88$, 95% CI [-1.19, -0.57]. In this case, however, group identity significantly moderated the magnitude of this blame reduction, $t(115) = 2.33$, $p = .02$. Specifically, learning about violations by outgroup members being unintentional reduced blame by 5.9 points less than learning about such violations by ingroup members. Consistent with our general predictions, however, blame change for both outgroup and ingroup members significantly differed from zero, ps $< .001$, $d = -0.82$ for ingroup targets, $d = -0.54$ for outgroup targets.

(2) In contrast to the previous studies, learning that a member of a group acted for morally bad reasons increased blame no more than learning merely that the group member acted intentionally ($M_{diff} = -0.52$), $F(1,115) = 0.59$, $p = .81$, $d = 0.03$, 95% CI [-0.21, 0.15]. This pattern was not moderated by the transgressor's ingroup or outgroup status, $F(1,115) = 0.13$, $p = .72$. Conversely, learning that a group member acted for morally good reasons substantially

[14] There were no significant differences in initial blame judgments for in-group targets ($M = 62.50$, $SD = 20.32$) versus outgroup targets ($M = 61.85$, $SD = 18.17$), $t(116) = 0.34$, $p = .74$.

reduced blame compared to intentionality alone (M_{diff} = -44.94), $F(1,115) = 201.24$, $p < .0001$, d = -1.78, 95% CI [-2.29, -1.27]. This mitigation was again not moderated by outgroup status, $F(1,115) = 2.34$, $p = .13$.

(3) Learning that a violation was preventable reduced blame less than unintentionality alone (M_{diff} = 6.48), $F(1,115) = 6.57$, $p = .012$, $d = 0.27$, 95% CI [0.01, 0.52]; this effect was not moderated by group status, $F(1,115) = 0.20$, $p = .66$. Learning that a violation was unpreventable reduced blame even further than unintentionality alone (M_{diff} = -11.98), $F(1,115) = 25.47$, $p < .001$, d = -0.46, 95% CI [-0.68, -0.24]; group status did not moderate the effect, $F(1,115) = 0.01$, $p = .94$.

Symmetry for in- and outgroup targets. We had found in the overall analyses that outgroup transgressors received more blame than ingroup transgressors, across information conditions. This tendency to exacerbate more and mitigate less should lead to violations of symmetry for outgroup transgressors. We first examined symmetry for ingroup transgressors. Exacerbation for intentional violations ($M = 23.49$, $SD = 26.02$) was indistinguishable from mitigation for unintentional violations ($M = 22.94$, $SD = 27.91$), $t(115) = 0.21$, $p = .83$, $d = 0.02$, 95% CI [-0.24, 0.28]; and exacerbation for morally bad reasons ($M = 22.28$, $SD = 26.81$) was indistinguishable from mitigation for morally good reasons ($M = 25.52$, $SD = 40.0$), $t(115) = -1.04$, $p = .30$, d = -0.09, 95% CI [-0.33, 0.14]. For outgroups, symmetry broke down. Exacerbation for intentional violations ($M = 25.64$, $SD = 26.49$) was larger than mitigation for unintentional violations ($M = 17.04$, $SD = 27.69$), $t(115) = 3.42$, $p < .001$, $d = 0.32$, 95% CI [0.05, 0.59]; and exacerbation for morally bad reasons ($M = 25.80$, $SD = 25.84$) was larger than mitigation for morally good reasons ($M = 15.23$, $SD = 41.46$), $t(115) = 3.38$, $p < .001$, $d = 0.31$, 95% CI [0.02, 0.60].

Discussion

In Study 6 a picture emerges in which moral perceivers blame outgroup members more overall (5 points on a 0-100 scale) than they blame ingroup members. This slightly harsher blame toward outgroup members breaks the updating symmetry consistently found in all five previous studies: for outgroup members, people mitigate less than they exacerbate blame (their initial blame judgments for the ambiguous first piece of information did not differ). However, even for outgroup members, people process new information in systematic and graded fashion as they do for ingroup members.

We can better understand these patterns of moral judgment for ingroup and outgroup members by comparing blame updating for these group targets with the average blame updating for individuals from Studies 1-5[15] (see Figure 8). This analysis reveals two key patterns. First, blame updating for ingroup targets never differs from that for individual targets (ps > .15). Second, bias against outgroup members is limited to more reluctant mitigation for the two conditions in which updating is maximally mitigating (i.e., good reasons, unpreventable accidents, ps < .022).

---

[15] We chose to use all five studies (not just the four on which our earlier meta-analysis was based) because Study 5 showed no differences between cognitive load and the standard condition, and the two together showed no notable differences from the average of Studies 1-4 (average difference in blame change means was 1.01, ranging from -2.50 to 3.37 across the six new information conditions). Further, we chose to conduct an analysis of mean comparisons rather than one of effect size comparisons because Study 6 used only half of the number of items per new information condition and therefore had considerably higher standard deviations than the previous five studies, making the scaling of effect sizes between Study 6 and Studies 1-5 incommensurable.

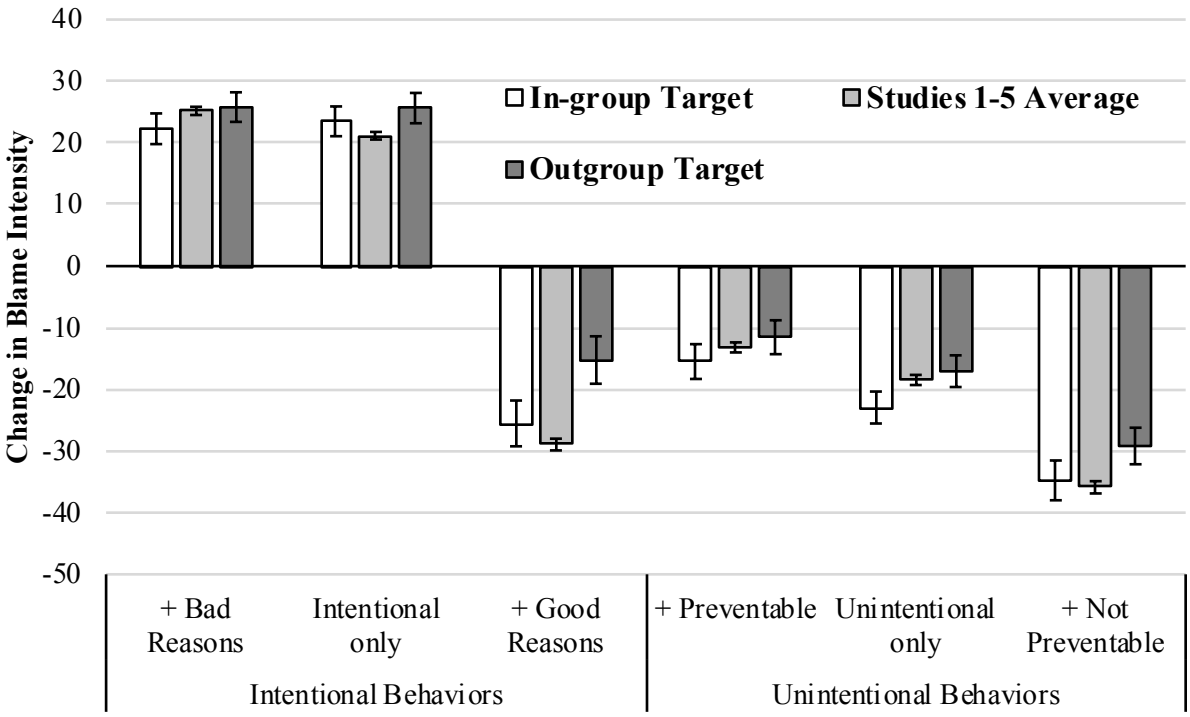

Figure 8. Targets' political group membership led to partially asymmetrical updating. People blamed political outgroup members more overall and in particular were less willing to mitigate for good reasons or unpreventable accidents. Error bars = ±1 SE.

These findings indicate that the social requirement for warrant is equally forceful for judgments of ingroup members as it is for everyday judgments where group membership is not salient. This result is broadly consistent with previous research showing that people are motivated to punish deviant ingroup members in order to teach moral rules or to preserve group cohesion (Boyd & Richerson, 1992; Cushman, 2013; Fehr & Gächter, 2002). Additionally, the data suggest that there is room for bias in blame judgments across group boundaries, but the pattern is not simply one of a stronger "desire to blame" (which should result in greater exacerbation). Instead, relaxing the requirement for warrant when judging outgroups may introduce the specter of bias by causing people to become actively skeptical of generally powerful mitigating information (e.g., an agent's good reasons or inability to prevent harm).

General Discussion

The socially-regulated perspective on blame suggests that because blame evolved as a form of costly social regulation, people are motivated to make blame judgments that would be considered fair and warranted by other moral perceivers in the social community. Thus, people systematically attend to evidence that can warrant a given blame judgment, including information about agents' mental states, their causal contributions to an outcome, and counterfactuals about the preventability of the outcome. Such systematic information processing should come into clear view when people have to update their blame judgments in response to new information. The present data support this hypothesis.

Studies using student samples (Study 1 and 5), community samples (Study 2), and Internet samples (Studies 3 to 4, 6) demonstrated consistent support for three gradedness predictions: (1) Relative to initial blame for a violation whose intentionality is ambiguous, people increased blame when they learned that a violation was intentional and decreased blame when they learned that a violation was unintentional. (2) Over and above changes due to intentionality, people increased blame when an agent acted for morally bad reasons and decreased blame substantially when an agent acted for morally good reasons. (3) Beyond changes after learning only that a violation was unintentional, people reduced blame less when an event was clearly preventable; however, when events were unpreventable people made even larger reductions in blame relative to unintentionality alone. These findings highlight the central role of causal and mental state information in the process of rendering blame judgments, and they are broadly consistent with previous research on single moral judgments (Cushman, 2008; Greene et al., 2009; Guglielmo & Malle, 2010; Lagnado & Channon, 2008; Monroe & Reeder, 2011; Young & Saxe, 2009).

Additionally, the present studies confirmed two symmetry predictions. The socially-regulated blame perspective makes the unique prediction that because blaming is a socially-costly act, perceivers should flexibly revise their blame judgments in response to any new relevant evidence, regardless of whether that evidence supports increasing or decreasing blame. The data bear out this claim. Examining the absolute magnitude of blame change showed that blame mitigation in response to learning that an agent *unintentionally* caused harm was equal to the blame exacerbation in response to learning that an agent *intentionally* caused harm. Similarly, blame mitigation in response to an agent's *morally good* reasons was equal to the blame exacerbation in response to an agent's *morally bad* reasons for acting.

Confirmation of the symmetry predictions is especially notable because they contrast with a key prediction of motivated-blame models. Such models predict that blame change should be asymmetrically biased against blame mitigation. Studies 1-5 showed little evidence supporting this bias prediction. The only exception to symmetry was in Study 6, when people made judgments of outgroup members. In this context, the community pressure to form fair and justified moral judgments is reduced and, indeed, we found that people's blame judgments were no longer symmetric and showed a bias toward overblaming outgroup members. Even so, mental state information had a strong impact on blame updating, accounting for nearly eight times as much variance in people's judgments ($\eta^2 = .55$) as did the ingroup/outgroup manipulation ($\eta^2 = .071$). Thus, bias co-existed with evidence-based information processing.

Theoretical Integration and Novel Predictions

The consistent pattern of graded and symmetric moral updating across a wide set of everyday moral infractions challenges the dominant view that moral information processing is routinely biased by early emerging moral judgments. However, these studies do not negate the possibility of motivational bias in blame. The social psychological literature is replete with examples of bias in social and moral judgments, and blame is no exception. Thus, although the studies here examine a situation where the socially-regulated blame and motivated-blame models make divergent predictions, we do not believe that these theories are, on the whole, mutually exclusive. Rather, the two theories are compatible insofar as they apply to different conditions of blaming.

When people are in a third-party role and consider everyday moral violations, when they have access to at least some causal-mental information, or when they make judgments across multiple different agents who are not obviously outgroup members, then their concern for

evidence-based blame judgments is activated and the Path Model of Blame successfully accounts for their judgments. But when people confront extreme acts of harm, make a single isolated judgment about an outgroup member, or when their judgments are anonymous and unchecked, motivated blame processes likely take hold. For example, in cases of personal injury to oneself or a loved one, people may not want to let the perpetrator "off the hook" (Fiske & Tetlock, 1997; Tetlock et al., 2007), and they may give less weight to intentionality or preventability and be more guided by their own desire to see the person be punished.

Central to the socially-regulated perspective is the claim that social demands for warrant motivate systematic moral information processing and judgments of blame. What follows from this contention is that intensifying or relaxing the requirement for warrant should modulate whether people are relatively more systematic or more biased in their judgments.

When the requirement for warrant is especially strong—for example, when blame is face-to-face, when social observers are present, or when one expects repeated interactions with the offender—perceivers are predicted to seek out morally relevant information and make graded, even-handed use of such information. When the requirement for warrant is relaxed—for example, when evaluating outgroup members about whom social observers care less, or when no social observers are present—perceivers are predicted to take short-cuts in seeking morally relevant information, become prone to asymmetric information processing, and render harsher blame judgments. A case in point are moral judgments expressed online (e.g., on Reddit, Twitter, or Facebook), where demands for warrant are paltry because people are anonymous, or address only ingroup members, and incur few costs for overblaming others. Crockett (2017) suggests that people express harsher moral condemnation online in part because of a supportive audience of like-minded others, a reduced risk of retaliation, and the ability to hide in a crowd.

Finally, the socially regulated perspective emphasizes a distinction between person-focused judgments of blame, for which warrant is necessary, and behavior-focused judgments of badness or wrongness, which are less costly and less socially-regulated (Voiklis & Malle, 2017). While people must be able to justify why the target of their blame judgment deserves the expressed amount of blame, calling an action wrong does not entail the same justificatory standard. Thus, whereas person-focused judgments of blame tend to be sensitive to causal-mental evidence and to the agreement or disagreement of social observers, behavior-focused judgments are less evidence-based, more susceptible to anchoring and asymmetrical adjustment, and less influenced by the presence or opinion of social observers.

Limitations and Future Directions for Research

Verbal stimuli and third-person perspective. A limitation of the present studies is that all of the stimuli are witness reports, verbally communicated to a third-party moral judge. This is the typical way in which theories of moral judgment are tested (including previous work on biases in blame), and it does constrain the generalizability of the results. In the present case, one might worry in particular about the minimalist initial reports about a violation (e.g., "Mark shot Frank"), which place a significant interpretational burden on perceivers. However, this burden is not unlike that put on consumers of daily news headlines (e.g., "Trump says to skip due process for those here illegally"; "Teenager missing after walking away from migrant center" – New York Times, 6/25/18). These headlines are designed to engage people, cause an evaluative response, and encourage them to seek further information, and this is just the situation our participants found themselves when asked to update their moral judgments. Nevertheless, we currently do not know how well these results match judgments arising from directly observed violations, when moral perceivers may be overcome with personal feelings of threat or outrage.

When perceivers are themselves victims, or when they are personally connected to the victim, blame judgments must deal with a unique emotional potential that could bias the moral processing (Patil, Calò, Fornasier, Cushman, & Silani, 2017) and lead to post-hoc justifications of preferred moral judgments (Alicke et al., 2011; Haidt, 2001).

The influence of emotion on moral updating. Research shows that manipulations of emotion can have a powerful effect on at least some types of moral judgment (Ask & Pina, 2011; Gutierrez & Giner-Sorolla, 2007; Monin, Pizarro, & Beer, 2007; Wheatley & Haidt, 2005). The present studies do not capture the full emotional dimension of moral judgments, and it seems unlikely that blame updating is immune to the effects of emotion. Exactly how emotion influences updated blame judgments is an important question for future research. One possibility is that negative emotional reactions (e.g., anger or disgust) amplify the intensity of people's moral judgments directly (Pizarro, Inbar, & Helion, 2011), creating harsher moral judgments overall without structurally changing the underlying information processing (about causality, intentionality, etc.). Alternatively, negative emotions may change information processing itself. Emotions may cause people to skip over critical pieces of evidence, adopt simple heuristics (e.g., "if you did it, then you're to blame"), or rely on pre-set values for morally relevant criteria (e.g., "these people always mean to hurt you"), resulting, for example, in an "intentionality bias" (Bègue, Bushman, Giancola, Subra, & Rosset, 2010; Rosset, 2008).

Conclusion

People make mistakes, in moral judgments as in nonmoral judgments. But we have learned, from these and other studies, that in many situations people make graded, sophisticated judgments. Two unique aspects of moral judgments make them perhaps more systematic than many others.

First, moral judgments are used for social regulation—to curtail and reform other people's behavior (Voiklis & Malle, 2017). If these judgments were often misguided or out of proportion, they would be ineffective at regulating others' behavior. In many respects modern society has succeeded remarkably in such regulation, enabling countless forms of coordination and cooperation among small groups and large crowds, and fostering civility to an extent that is arguably unsurpassed in human history (Pinker, 2011). The success of this social regulation attests to the success of moral judgment to support such regulation.

Second, moral judgments are themselves under social regulation. Because moral judgments, when socially expressed, can hurt or harm the target, there are norms in place to demand the judgments to be fair and proportional. A community that condones outrageous and exaggerated blame and punishment will not succeed, because such unfair moral criticism leads to anger, retaliation, and perceptions of injustice. Successful communities therefore demand of their members to render reasonably thoughtful and evidence-based moral judgments, especially blame judgments, which are directed at persons and therefore most costly when unfounded.

Many domains of psychology have moved away from merely recounting the errors that humans commit and have turned instead to investigating bounded rationality and relatively calibrated judgments under uncertainty. This is surely not to claim perfection of the human mind. But scientists might do well to respect and admire the achievements of the human mind at least as much as its downfalls.

References

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

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*, 556–574. doi:10.1037/0033-2909.126.4.556

Alicke, M. D., Rose, D., & Bloom, D. (2011). Causation, norm violation and culpable control. *Journal of Philosophy*, *108*, 670–696.

Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science*. doi:10.1177/0956797613480507

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*, 1423–1440. doi:10.1037/0022-3514.83.6.1423

Ask, K., & Pina, A. (2011). On being angry and punitive: How anger alters perception of criminal intent. *Social Psychological and Personality Science*, *2*, 494–499. doi:10.1177/1948550611398415

Barrett, H. C., Todd, P. M., Miller, G. F., & Blythe, P. W. (2005). Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior*, *26*, 313–331. doi:10.1016/j.evolhumbehav.2004.08.015

Bègue, L., Bushman, B. J., Giancola, P. R., Subra, B., & Rosset, E. (2010). "There is no such thing as an accident," especially when people are drunk. *Personality and Social Psychology Bulletin*, *36*, 1301–1304. doi:10.1177/0146167210383044

Bergmann, J. R. (1998). Introduction: Morality in discourse. *Research on Language & Social Interaction*, *31*, 279–294.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. New York, NY: Cambridge University Press.

Boehm, C. (2000). The origin of morality as social control. *Journal of Consciousness Studies*, *7*, 149–183.

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, *13*, 171–195. doi:10.1016/0162-3095(92)90032-Y

Carnes, N. C., Lickel, B., & Janoff-Bulman, R. (2015). Shared perceptions: Morality is embedded in social contexts. *Personality and Social Psychology Bulletin*, 0146167214566187. doi:10.1177/0146167214566187

Chernyak, N., & Sobel, D. M. (2016). "But he didn't mean to do it": Preschoolers correct punishments imposed on accidental transgressors. *Cognitive Development*, *39*, 13–20. doi:10.1016/j.cogdev.2016.03.002

Coates, D. J., & Tognazzini, N. A. (2012). The contours of blame. In D. J. Coates & N. A. Tognazzini (Eds.), *Blame: Its nature and norms* (pp. 3–26). New York, NY: Oxford University Press.

Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*, 37–57. doi:10.1037/pspa0000014

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. In J. M. Olson (Ed.), *Advances in Experimental Social Psychology* (Vol. 56, pp. 131–199). San Diego, CA, US: Elsevier.

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. (2013). The functional design of punishment and the psychology of learning. In R. Joyce, K. Sterelny, B. Calcott, & B. Fraser (Eds.), *Psychological and environmental foundations of cooperation.*, Signaling, commitment and emotion (Vol. 2). Cambridge, MA: MIT Press.

Decety, J., & Cacioppo, S. (2012). The speed of morality: a high-density electrical neuroimaging study. *Journal of Neurophysiology*, *108*, 3068–3072. doi:10.1152/jn.00473.2012

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*, 568–584.

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Moral judgment and decision making.*, The psychology of learning and motivation; Vol 50; 0079-7421 (Print); (Vol. 50, pp. 307–338). San Diego, CA US: Elsevier Academic Press.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140. doi:10.1038/415137a

Finkel, N. J., Maloney, S. T., Valbuena, M. Z., & Groscup, J. L. (1995). Lay perspectives on legal conundrums. *Law and Human Behavior*, *19*, 593–608. doi:10.1007/BF01499376

Fiske, A. P., & Tetlock, P. E. (1997). Taboo Trade-offs: Reactions to Transactions That Transgress the Spheres of Justice. *Political Psychology*, *18*, 255–297. doi:10.1111/0162-895X.00058

Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the Message Punishment Is Satisfying If the Transgressor Responds to Its Communicative Intent. *Personality and Social Psychology Bulletin*, 0146167214533130. doi:10.1177/0146167214533130

Gardner, T & Monroe, A. E. (2018). *Blaming the blamers: People impose different obligations to punish on victims versus third parties.* Poster presented at the Society of Southeastern Social Psychologists Conference, Raleigh, NC.

Gilovich, T. (1983). Biased evaluation and persistence in gambling. *Journal of Personality and Social Psychology*, *44*, 1110–1126.

Gollwitzer, M., & Denzler, M. (2009). What makes revenge sweet: Seeing the offender suffer or delivering a message? *Journal of Experimental Social Psychology*, *45*, 840–844. doi:10.1016/j.jesp.2009.03.001

Gollwitzer, M., Meder, M., & Schmitt, M. (2011). What gives victims satisfaction when they seek revenge? *European Journal of Social Psychology*, *41*, 364–374. doi:10.1002/ejsp.782

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148–168. doi:10.1037/a0034726

Gray, K., & Wegner, D. M. (2008). The sting of intentional pain. *Psychological Science*, *19*, 1260–1262. doi:10.1111/j.1467-9280.2008.02208.x

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

Greene, J. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. New York: Penguin Press.

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral*

psychology, Vol 3: The neuroscience of morality: Emotion, brain disorders, and development. (pp. 35–80). Cambridge, MA: MIT Press.

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*, 364–371. doi:10.1016/j.cognition.2009.02.001

Guala, F. (2012). Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. *The Behavioral and Brain Sciences*, *35*, 1–15. doi:10.1017/S0140525X11000069

Guglielmo, S. (2015). Moral judgment as information processing: an integrative review. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.01637

Guglielmo, S., & Malle, B. F. (2010). Can unintended side effects be intentional? Resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin*, *36*, 1635–1647. doi:10.1177/0146167210386733

Guglielmo, S., Monroe, A. E., & Malle, B. F. (2009). At the heart of morality lies folk psychology. *Inquiry*, *52*, 449–466. doi:10.1080/00201740903302600

Gutierrez, R., & Giner-Sorolla, R. (2007). Anger, disgust, and presumption of harm as reactions to taboo-breaking behaviors. *Emotion*, *7*, 853–868. doi:10.1037/1528-3542.7.4.853

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

Haidt, J. (2007). The New Synthesis in Moral Psychology. *Science*, *316*, 998–1002. doi:10.1126/science.1137651

Haidt, J. (2008). Morality. *Perspectives on Psychological Science*, *3*, 65–72.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. doi:10.1017/S0140525X0999152X

Hulsizer, M. R., Munro, G. D., Fagerlin, A., & Taylor, S. P. (2004). Molding the past: Biased assimilation of historical information. *Journal of Applied Social Psychology*, *34*, 1048–1074. doi:10.1111/j.1559-1816.2004.tb02583.x

Ingram, G. (2014). From hitting to tattling to gossip: An evolutionary rationale for the development of indirect aggression. *Evolutionary Psychology*, *12*, 343–363.

Kammrath, L. K., Ames, D. R., & Scholer, A. A. (2007). Keeping up impressions: Inferential rules for impression change across the Big Five. *Journal of Experimental Social Psychology*, *43*, 450–457. doi:10.1016/j.jesp.2006.04.006

Kim, B., Voiklis, J., Cusimano, C., & Malle, B. F. (2015, February). *Norms of moral criticism: Do people prohibit underblaming and overblaming?* Poster presented at the Annual meeting of the Society of Personality and Social Psychology, Long Beach, CA.

Knauft, B. M. (1991). Violence and sociality in human evolution. *Current Anthropology*, *32*, 391–428.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*, 190–194. doi:10.1111/1467-8284.00419

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*, 754–770. doi:10.1016/j.cognition.2008.06.009

Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin*, *24*, 563–574.

Lieberman, D., & Linke, L. (2007). The effect of social category on third party punishment.

Evolutionary Psychology, *5*, 289–305.
Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., & Blair, R. J. R. (2006). The neural basis of implicit moral attitude—An IAT study using event-related fMRI. *NeuroImage*, *30*, 1449–1457. doi:10.1016/j.neuroimage.2005.11.005
MacCoun, R. J. (2005). Voice, control, and belonging: The double-edged sword of procedural fairness. *Annual Review of Law and Social Science*, *1*, 171–201. doi:10.1146/annurev.lawsocsci.1.041604.115958
Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, *3*, 23–48. doi:10.1207/s15327957pspr0301_2
Malle, B. F., Guglielmo, S., & Monroe, A. E. (2012). Moral, cognitive, and social: The nature of blame. In J. P. Forgas, K. Fiedler, & C. Sedikides (Eds.), *Social Thinking and Interpersonal Behavior*, Sydney symposium of social psychology (pp. 313–331). New York, NY US: Psychology Press.
Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, *25*, 147–186. doi:10.1080/1047840X.2014.877340
Malle, B. F., & Holbrook, J. (2012). Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality and Social Psychology*, *102*, 661–684. doi:10.1037/a0026790
Martin, J. W., & Cushman, F. (2016). Why we forgive what can't be controlled. *Cognition*, *147*, 133–143. doi:10.1016/j.cognition.2015.11.008
Mazzocco, P. J., Alicke, M. D., & Davis, T. L. (2004). On the robustness of outcome bias: No constraint by prior culpability. *Basic and Applied Social Psychology*, *26*, 131–146. doi:10.1080/01973533.2004.9646401
Mende-Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic Value Underlies Asymmetric Updating of Impressions in the Morality and Ability Domains. *The Journal of Neuroscience*, *33*, 19406–19415. doi:10.1523/JNEUROSCI.2334-13.2013
Mende-Siedlecki, P., Cai, Y., & Todorov, A. (2012). The neural dynamics of updating person impressions. *Social Cognitive and Affective Neuroscience*. doi:10.1093/scan/nss040
Mende-Siedlecki, P., & Todorov, A. (2016). Neural dissociations between meaningful and mere inconsistency in impression updating. *Social Cognitive and Affective Neuroscience*, 1489–1500. doi:10.1093/scan/nsw058
Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, *11*, 143–152. doi:10.1016/j.tics.2006.12.007
Mikula, G., Petri, B., & Tanzer, N. (1990). What people regard as unjust: Types and structures of everyday experiences of injustice. *European Journal of Social Psychology*, *20*, 133–149. doi:10.1002/ejsp.2420200205
Mikula, G., Scherer, K. R., & Athenstaedt, U. (1998). The Role of Injustice in the Elicitation of Differential Emotional Reactions. *Personality and Social Psychology Bulletin*, *24*, 769–783. doi:10.1177/0146167298247009
Miller, D. T. (2001). Disrespect and the Experience of Injustice. *Annual Review of Psychology*, *52*, 527–553. doi:10.1146/annurev.psych.52.1.527
Monin, B., Pizarro, D. A., & Beer, J. S. (2007). Deciding versus reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology*, *11*, 99–111. doi:10.1037/1089-2680.11.2.99
Monroe, A. E., & Malle, B. F. (2017). Two paths to blame: Intentionality directs moral information processing along two distinct tracks. *Journal of Experimental Psychology:*

General, *146*, 123–133.
Monroe, A. E., & Reeder, G. D. (2011). Motive-matching: Perceptions of intentionality for coerced action. *Journal of Experimental Social Psychology*, *47*, 1255–1261. doi:10.1016/j.jesp.2011.05.012
Munro, G. D., Weih, C., & Tsai, J. (2010). Motivated suspicion: Asymmetrical attributions of the behavior of political ingroup and outgroup members. *Basic and Applied Social Psychology*, *32*, 173–184. doi:10.1080/01973531003738551
Nadelhoffer, T. (2006). Bad acts, blameworthy agents, and intentional actions: Some problems for juror impartiality. *Philosophical Explorations*, *9*, 203–219.
Nadler, J. (2012). Blaming as a social process: The influence of character and moral emotion on blame. *Law and Contemporary Problems*, *75*, 1–31.
Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding mechanical turk as a participant pool. *Current Directions in Psychological Science*, *23*, 184–188. doi:10.1177/0963721414531598
Patil, I., Calò, M., Fornasier, F., Cushman, F., & Silani, G. (2017). The behavioral and neural basis of empathic blame. *Scientific Reports*.
Pettit, D., & Knobe, J. (2009). The pervasive impact of moral judgment. *Mind and Language*, *24*, 586–604. doi:10.1111/j.1468-0017.2009.01375.x
Pinker, S. (2011). *The better angels of our nature: why violence has declined*. New York, NY: Viking.
Pizarro, D. A., & Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil.*, Herzliya series on personality and social psychology (pp. 91–108). Washington, DC US: American Psychological Association.
Pizarro, D., Inbar, Y., & Helion, C. (2011). On disgust and moral judgment. *Emotion Review*, *3*, 267–268. doi:10.1177/1754073911402394
Plaks, J. E., McNichols, N. K., & Fortune, J. L. (2009). Thoughts versus deeds: Distal and proximal intent in lay judgments of moral responsibility. *Personality and Social Psychology Bulletin*, *35*, 1687–1701. doi:10.1177/0146167209345529
Przepiorka, W., & Berger, J. (2016). The Sanctioning Dilemma: A Quasi-Experiment on Social Norm Enforcement in the Train. *European Sociological Review*, *32*, 439–451. doi:10.1093/esr/jcw014
Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*, 57–75. doi:10.1037/a0021867
Rapp, D. N., & Kendeou, P. (2007). Revising what readers know: Updating text representations during narrative comprehension. *Memory & Cognition*, *35*, 2019–2032. doi:10.3758/BF03192934
Reeder, G., D., & Brewer, M., B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, *86*, 61–79. doi:10.1037/0033-295x.86.1.61
Reeder, G., D., & Coovert, M., D. (1986). Revising an impression of morality. *Social Cognition*, *4*, 1–17. doi:10.1521/soco.1986.4.1.1
Reeder, G. D., Monroe, A. E., & Pryor, J. B. (2008). Impressions of Milgram's obedient teachers: Situational cues inform inferences about motives and traits. *Journal of*

Personality and Social Psychology, *95*, 1–17. doi:10.1037/0022-3514.95.1.1
Reeder, G. D., Pryor, J. B., Wohl, M. J. A., & Griswell, M. L. (2005). On Attributing Negative Motives to Others Who Disagree With Our Opinions. *Personality and Social Psychology Bulletin*, *31*, 1498–1510. doi:10.1177/0146167205277093
Robinson, P. H., & Darley, J. M. (1995). *Justice, liability, and blame: Community views and the criminal law*. New directions in social psychology. Boulder, CO: Westview Press.
Rosset, E. (2008). It's no accident: Our bias for intentional explanations. *Cognition*, *108*, 771–780. doi:10.1016/j.cognition.2008.07.001
Schiller, B., Baumgartner, T., & Knoch, D. (2014). Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination. *Evolution and Human Behavior*. doi:10.1016/j.evolhumbehav.2013.12.006
Scholten, L., van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. W. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology*, *43*, 539–552. doi:10.1016/j.jesp.2006.05.010
Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness.* New York, NY US: Springer Verlag.
Silberbauer, G. (1982). Political process G/wi bands. In E. Leacock & R. Lee (Eds.), *Politics and history in band societies* (pp. 23–36). Cambridge, England: Cambridge University Press.
Strohminger, N., & Nichols, S. (2015). Neurodegeneration and Identity. *Psychological Science*, 0956797615592381. doi:10.1177/0956797615592381
Stuyven, E., Van der Goten, K., Vandierendonck, A., Claeys, K., & Crevits, L. (2000). The effect of cognitive load on saccadic eye movements. *Acta Psychologica*, *104*, 69–85. doi:10.1016/S0001-6918(99)00054-2
Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*, 96–102.
Tetlock, P. E. (1985). Accountability: A Social Check on the Fundamental Attribution Error. *Social Psychology Quarterly*, *48*, 227–236. doi:10.2307/3033683
Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. *Journal of Personality and Social Psychology*, *57*, 632–640. doi:10.1037//0022-3514.57.4.632
Tetlock, P. E., Visser, P. S., Singh, R., Polifroni, M., Scott, A., Elson, S. B., Mazzocco, P., et al. (2007). People as intuitive prosecutors: The impact of social-control goals on attributions of responsibility. *Journal of Experimental Social Psychology*, *43*, 195–209. doi:10.1016/j.jesp.2006.02.009
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
Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological science*, *20*, 1092–1099. doi:10.1111/j.1467-9280.2009.02411.x
Voiklis, J., & Malle, B. F. (2017). Moral cognition and its basis in social cognition and social regulation. In K. Gray & J. Graham (Eds.), *Atlas of Moral Psychology* (pp. 108–120). New York, NY: Guilford.
Wallace, R. J. (1994). *responsibility and the moral sentiments*. Cambridge, MA: Harvard University Press.
Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*.

Judgments of responsibility: A foundation for a theory of social conduct. New York, NY US: Guilford Press.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, *16*, 780–784. doi:10.1111/j.1467-9280.2005.01614.x

Wilson, D. S. (2010). *Darwin's cathedral: Evolution, religion, and the nature of society*. University of Chicago Press.

Wilson, P. J. (1991). *The domestication of the human species*. New Haven, CT: Yale University Press.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the Dominance of Moral Categories in Impression Formation. *Personality and Social Psychology Bulletin*, *24*, 1251–1263. doi:10.1177/01461672982412001

Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, *100*, 283–301. doi:10.1016/j.cognition.2005.05.002

Young, L., & Saxe, R. (2009). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia*, *47*, 2065–2072. doi:10.1016/j.neuropsychologia.2009.03.020