

---

## TARGET ARTICLE

---

### A Theory of Blame

**Bertram F. Malle**

*Department of Cognitive, Linguistic, and Psychological Sciences, Brown University,  
Providence, Rhode Island*

**Steve Guglielmo**

*Department of Psychology, Macalester College, Saint Paul, Minnesota*

**Andrew E. Monroe**

*Department of Psychology, Florida State University, Tallahassee, Florida*

We introduce a theory of blame in five parts. Part 1 addresses what blame is: a unique moral judgment that is both cognitive and social, regulates social behavior, fundamentally relies on social cognition, and requires warrant. Using these properties, we distinguish blame from such phenomena as anger, event evaluation, and wrongness judgments. Part 2 offers the heart of the theory: the Path Model of Blame, which identifies the conceptual structure in which blame judgments are embedded and the information processing that generates such judgments. After reviewing evidence for the Path Model, we contrast it with alternative models of blame and moral judgment (Part 3) and use it to account for a number of challenging findings in the literature (Part 4). Part 5 moves from blame as a cognitive judgment to blame as a social act. We situate social blame in the larger family of moral criticism, highlight its communicative nature, and discuss the darker sides of moral criticism. Finally, we show how the Path Model of Blame can bring order to numerous tools of blame management, including denial, justification, and excuse.

*Key words:* morality, responsibility, social cognition, intentionality, judgment, emotion

For centuries, “moral psychology” referred to a domain of inquiry in philosophical ethics. Over the past decade, however, a substantial body of theoretical and empirical work has emerged that constitutes “moral psychology” as an interdisciplinary field poised to answer fundamental questions about mind and sociality: How do norms and values guide behavior? What faculties underlie moral judgment and moral action? How do these faculties relate to social cognition and emotion?

Our goal in this article is to elucidate one central element of moral psychology: blame. Blame, wrote Beardsley (1970), “has a power and poignancy for human life unparalleled by other moral concepts” (p. 176). We introduce a theory of blame in five parts. Part 1 addresses what blame is and is not. We propose that it is a unique type of moral judgment and has four

properties: It is both cognitive and social; it regulates social behavior; it fundamentally relies on social cognition; and, as a social act, it requires warrant. These four properties allow us to distinguish blame from several other phenomena, such as anger, event evaluation, and wrongness judgments.

Part 2 offers the heart of the theory: the Path Model of Blame, which identifies the conceptual structure in which blame judgments are embedded and the information processing that generates such judgments. We also review the substantial indirect and more recent direct evidence for the Path Model of Blame.

Part 3 contrasts the Path Model with a number of alternative models of blame and moral judgment, including responsibility models, models of motivated blame, and models of affect-based moral judgment.

Part 4 introduces a number of challenging findings in the moral psychology literature and probes how the Path Model can account for them.

Part 5 moves from blame as a cognitive judgment to blame as a social act. We situate social blame in the larger family of moral criticism, highlight its communicative nature and constructive potential, but also discuss the darker sides of moral criticism. Finally, we show how the Path Model of Blame can bring order to numerous findings on social blame management, including denial, justification, and excuse.

### Three Types of Moral Judgment

In the family of moral judgments we must distinguish at least three types<sup>1</sup>:

1. Setting and affirming *norms*, such as declaring a prohibition, expressing an imperative, or avowing one norm as overriding another.
2. Evaluating *events* (outcomes, behaviors) in light of those norms, such as by judging an event as *bad*, *good*, *wrong*, or (*im*)*permissible*.
3. Evaluating *agents* for their involvement in such norm-relevant events, such as by judging someone as *morally responsible*, *blameworthy*, or *praiseworthy*.

The key difference between these three types of judgment is that Type 1 engages directly with norms, whereas Types 2 and 3 make evaluative judgments in light of those norms, with Type 2 directed at events and Type 3 directed at agents. We mostly set aside Type 1 judgments and assume that moral perceivers have some norm system (Nichols, 2002) but sometimes vehemently disagree over specific norms (Skitka, Bauman, & Sargis, 2005; Tetlock, 2003). We focus on blame as the paradigmatic Type 3 judgment but show how it both relies on and goes beyond Type 2 judgments.

### Part 1: What Blame Is and Is Not

#### What Blame Is: Four Fundamental Properties

##### 1. *Blame Is Cognitive and Social*

The cognitive, private side of blame is the process that leads to a judgment of blame; the social, public side is the act of expressing a blame judgment to another person. When and why cognitive blame occurs (e.g., in response to certain stimuli, with

characteristic information processing, aided by certain emotions) differs from when and why social blame occurs (e.g., guided by goals, roles, and norms). A comprehensive theory of blame must address both sides, as well as the relationship between them (Coates & Tognazzini, 2012a). This relationship is typically described in only one direction, as social blame expressing cognitive blame (Beardsley, 1970; Zaibert, 2005). But we propose that the relationship also goes in the other direction: that cognitive blame is critically constrained by and inherits properties from social blame.

##### 2. *Blame Is Social Regulation*

Morality regulates individual behaviors so they come in line with community interests and sustain social relations (Deigh, 1996; Flack & de Waal, 2000; Haidt, 2008; Joyce, 2006; Rai & Fiske, 2011). Part of this morality rests on biological foundations in mammal social-emotional life (Churchland, 2012; de Waal, 2006). Those include motives for belonging, caring, and shared experience. But in human history, biological instincts alone did not suffice for social regulation. People had to be motivated to act not only in accordance with their intrinsic social desires (e.g., to belong, to be accepted; Baumeister & Leary, 1995) but also in accordance with social expectations for sharing (e.g., food), reciprocity, self-control (e.g., politeness, modesty), and recognition of others' rights and vulnerabilities. This kind of cultural morality regulates behavior by way of norms and values (Sripada & Stich, 2006; Sunstein, 1996; Thierry, 2000), which have been taught, learned, and enforced during humans' nomadic small-group past (Wiessner, 2005; Woodburn, 1982) and were vastly expanded in the last 10,000 years (Tiger, 2000). Of importance, cultural morality has succeeded by tying norm compliance to the fulfillment of social-biological needs: adhering to norms promises positive social relations, status, resources, and shared experiences, whereas violating norms jeopardizes these social benefits (Chudek & Henrich, 2011). Blaming and praising people for their behaviors is a key mechanism to implement such patterns of social-cultural regulation (Cushman, 2013).

##### 3. *Blame Relies on Social Cognition*

Because blame's primary and original function is to publicly regulate community members' conduct, it is a judgment directed at a *person* who has caused or done something norm violating (e.g., Scanlon, 2008; Sher, 2006). As a person judgment, blame relies on person perception or "social cognition"—the suite of concepts and processes that allow people to make sense of human behavior (Malle, 2008). Social cognitive information processing comes for free, as it

<sup>1</sup>A potential fourth type comprises ascriptions of moral dispositions, but nothing in our theory depends on whether such a distinct fourth type exists.

were, for judgments of blame (Guglielmo, Monroe, & Malle, 2009). Of importance, a subset of this social-cognitive information serves as conditions or “criteria” for assigning blame, most prominently intentionality and mental states (Alicke, 2000; Cushman, 2008; Guglielmo et al., 2009; Shaver, 1985). These particular social-cognitive criteria underlie blame, we suspect, because of their effectiveness in regulating behavior (McGeer, 2012a, 2012b). For example, by strongly responding to intentional norm violations and by blaming preventable but not unpreventable unintentional behaviors, moral perceivers focus on the behaviors that are most under the agent’s control.

#### 4. Blame Requires Warrant

Because social blame regulates behavior by criticizing or even devaluing the blamed agent, it is a strong and potentially damaging intervention. As a result, acts of blaming are themselves subject to social norms (Coates & Tognazzini, 2012b). In particular, social blaming carries a burden of *warrant*: The blamer must be able to offer grounds for why the agent deserves the attributed blame (McKenna, 2012). Whereas one can say, “It’s just wrong, I can’t tell you why,” it would be socially unacceptable to say, “He deserves blame, but I can’t tell you why.”<sup>2</sup> One of the pivotal ways in which social blame and cognitive blame are intertwined is that the warrant for social blame resides in large part in the very criteria on which people normally base their cognitive judgments of blame (Roskies & Malle, 2013), such as causality, intentionality, and preventability. (We discuss these criteria in detail in the next section.) Because of this demand of warrant for social blame, the blamer must not only acquire information that counts as such warrant but also keep this information accessible when expressing a judgment of blame. And even though the blamer can be in error, can confabulate or lie, the community can fact-check the blamer’s warrant. We suggest that one of the major properties of blame is that the demand on social blame to offer warrant puts pressure on the fidelity and transparency of cognitive blame (cf. Lerner & Tetlock, 1999).

<sup>2</sup>Findings on “moral dumbfounding” (Haidt, Koller, & Dias, 1993; Uhlmann & Zhu, 2013) all concern Type 2 judgments of why something is “wrong,” not Type 3 judgment such as why somebody is to blame. It may be true that people often do not have access to the origins and justifications of their event-directed (Type 2) moral judgments. But things are different for Type 3 judgments like blame: Only if people generally have access to the informational basis of their blame judgments (e.g., inferences of intentionality or preventability) can they demand, offer, and negotiate such information as warrants for their acts of blaming. Indeed, Bucciarelli, Khemlani, and Johnson-Laird (2008, Study 3) showed that people had no trouble explicating (in a think-aloud protocol) why one of two agents was more blame- or praiseworthy than another.

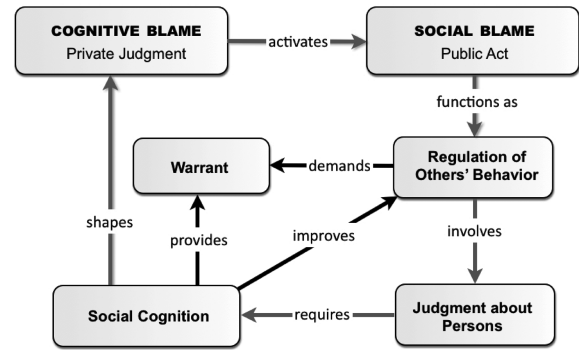


Figure 1. Relationships between cognitive and social blame. (Color figure available online.)

We depict the relationships among the social and cognitive properties of blame in Figure 1. Having proposed what blame is, we can proceed to state what blame is not.

#### What Blame Is Not

##### Blame Is Not Merely Anger

Blame judgments and social acts of blame are frequently (but not necessarily) accompanied by anger. Anger and blame share some properties (e.g., both are easily elicited by injustice; Wranik & Scherer, 2010), and some researchers even characterize anger as *relying* on attributions of blame (e.g., Averill, 1983), but the two should not be equated (Berkowitz & Harmon-Jones, 2004). There is the nontrivial fact that we can say, “He felt anger” but not “He felt blame.” There are cases of blaming without anger (e.g., participants in experiments who make blame ratings about fictitious behaviors; people with high levels of patience or compassion; Pettigrove & Tanaka, 2013); and there are cases of anger without blaming (K. B. Anderson, Anderson, Dill, & Deuser, 1998; Herrald & Tomaka, 2002). More systematically, anger differs on several of blame’s defining properties: Unlike blame, anger can be directed at or caused by impersonal events (e.g., unpleasant weather, C. A. Anderson, Deuser, & DeNeve, 1995; physical pain, Fernandez & Turk, 1995); anger can and often does occur without accessible warrant (“I am just angry at her, I don’t know why”; cf. Shaver, Schwartz, Kirson, & O’Connor, 1987); and, by itself, anger is not an effective tool of social regulation.<sup>3</sup>

<sup>3</sup>There has been some debate in philosophy about whether anger may have positive social functions, but these proposals discuss not mere anger but “angry blame” (Wolf, 2011), “righteous blame” (Frye, 1983), “moral anger” (Prinz, 2007) or indignation and resentment (MacLachlan, 2010)—all suspiciously close to blame itself and therefore likely to inherit the social functions of blame.

### ***Blame Is Not Merely Event Evaluation***

According to Haidt (2001), “Moral judgments are . . . defined as evaluations (good versus bad) of the actions or character of a person” (p. 817). We agree that people often make such good–bad evaluations, both about nonbehavioral events (a broken window) and behavioral events (a person breaking a window). But these are what we have called Type 2 moral judgments, lacking all of blame’s properties: they are not about a person; they rarely require social-cognitive information (e.g., intentionality, reasons), they do not demand warrant, and they only indirectly regulate behavior by reaffirming a norm.

### ***Blame Is Not Merely a Wrongness Judgment***

When examining lay definitions of blame, Pearce (2003) found that fewer than 2% of definitions referred to the wrongness of a behavior, and Cushman (2008) showed that people differentiate between wrongness and blame. Within our theoretical framework, too, several properties distinguish blame from wrongness judgments.

First, whereas blame judgments target an agent, wrongness judgments target a behavior, and typically an intentional one (“stealing is wrong”; “it was wrong not to tell her the truth”). A participant in Haidt and Hersch’s (2001, p. 210) study illustrates the distinction between these judgments. When explaining why she objected to gay male intercourse, she said, “I don’t think it’s their fault, I don’t blame them, but I still, I, I have a problem, morally with it.” She does not blame the *persons* for engaging in the behavior, but she finds the *behavior* morally wrong.

Second, as mentioned earlier, whereas blame judgments require warrant, wrongness judgments do not. When saying something is wrong, people often simply assert *that* a norm has been violated: “It’s just morally wrong!” (*CBS Evening News*, April 25, 2010) and explicate at most *which* norm was violated: “What James had done was wrong because it violated pre-existing rights of Englishmen” (Claus, 2004, p. 136); “war is wrong because it conflicts with Christian principles” (Watson, 1999, p. 64). In sharp contrast, blame judgments are warranted by citing information specific to the *person* committing the norm violation, such as causality (“her parents were to blame for her obesity because they’d started over-feeding her at birth”; Morrison, 2010, p. 14), capacity (“I blame the police department because . . . they could have nipped this in the bud”; Rivera, August 19, 1992), obligation (“He should have tried . . . to get her some help”; Hogan, April 10, 2007); and above all, mental states (e.g., “The chairman knew that his action would have caused damage”; “He did not really care about the environment”; Zalla & Leboyer, 2011).

We summarize in Table 1 the properties of blame and how these properties distinguish blame from other judgments.

With this understanding of what blame is and is not, we turn to the concepts and information processing that underlie cognitive blame judgments and that provide warrant for social blame. We should emphasize that this focus on concepts and information processing in no way denies the role of affect and emotion in blame or the possibility of motivated reasoning. In fact, because our model identifies the

Table 1. *Properties of Blame and How They Distinguish Blame From Related Constructs.*

	Directed at What Object	Relying on Social Cognition?	Social Regulation of Behavior?	Warrant?
<b>Blame judgment</b>	Persons	Yes: intentionality, mental states	Direct by way of public criticism	Yes: by citing person information
<b>Wrongness judgment</b>	Actions	Partial: coding for intentionality	Direct when calling out person’s action; indirect when affirming norm	No: declaring <i>that</i> a norm was violated
<b>Anger</b>	Anything (persons, behaviors, outcomes)	Sometimes: if directed at a person’s motives	Variable	No: citing only cause of anger
<b>Event evaluation</b>	Events	Minimal	Indirect by affirming norm	No: mere statement of event valence

specific information processing components that give rise to blame judgments we are able to pinpoint, in a later section, more precisely the involvement of affect, emotion, and motivation. But we must first fully capture the complexity of information processing underlying blame.

## Part 2: The Path Model of Blame

### Overview

The model posits that blame judgments arise within a conceptual structure already in place in ordinary social cognition, involving concepts such as *cause*, *agent*, *intentionality*, and *reasons*. Blame judgments therefore rely on familiar psychological processes operating over these concepts (Malle, 2005, 2008), including causal reasoning, intentionality judgments, and mental state inferences. But in service of generating a blame judgment, these concepts and processes follow a logic of *criteria*. As posited earlier, social acts of blame can be costly and require warrant, and the cognitive judgments that underlie such acts of blame are constrained by this requirement. Blame judgments therefore involve integrating information relevant to certain critical concepts and “testing” whether the criteria are met. A cognitive system can either test a given set of criteria simultaneously to deliver the relevant judgment (Alicke, 2000; N. H. Anderson, 1991; Schlenker, Britt, Pennington, Murphy, & Doherty, 1994) or rely on a nested logic such that certain criteria are generally tested first and, depending on their value, processing of subsequent criteria is omitted, engaged, or terminated. Processing en route to blame, we propose, exploits such a nested logic by proceeding along particular paths, which are represented by the ordered structure in Figure 2.

Within this structure, blame emerges if the social perceiver

- detects that an event or outcome<sup>4</sup> violated a norm; and
- determines that an *agent caused* the event.

If no agent (person or group) is causally linked to the norm violation, the social perceiver may feel angry, sad, or worried, but blame does not arise because there is not target for it. If agent causality is established, however, the perceiver

- *judges whether the agent brought about the event intentionally.*

<sup>4</sup>Events are time-extended processes (e.g., a car skidding on ice; a person firing a gun at someone), whereas outcomes are the results of events (e.g., a damaged car; a dead person). Our model applies equally well to both.

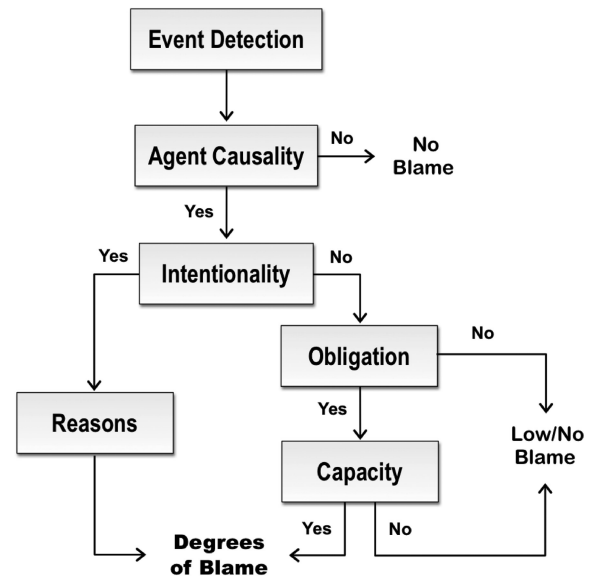


Figure 2. Concepts and processing paths in the Path Model of Blame. Note. Obligation = obligation to prevent the event in question; Capacity = capacity to prevent the event in question.

Once this judgment is made, two very different information-processing paths lead to blame.

If the agent is judged to have acted intentionally, the perceiver

- considers the agent’s *reasons* for acting.

Blame is then graded depending on the justification these reasons provide—minimal blame if the agent was justified in acting this way; maximal blame if the agent was not justified.

If the agent is judged to have brought about the event unintentionally, the perceiver

- considers whether the agent *should* have prevented the norm-violating event (*obligation*) and
- considers whether the agent *could* have prevented the event (*capacity*).

### Clarifications

We offer three points of clarification. First, there is no restriction built into the Path Model regarding the modes of processing (e.g., automatic vs. controlled, conscious vs. unconscious) by which moral perceivers arrive at a blame judgment. Any given component’s appraisal (e.g., about agentic causality or intentionality) may in principle be automatic or controlled, conscious or unconscious, depending on such factors as stimulus salience, existing knowledge structures, cognitive load, and so on (Kruglanski & Orehek, 2007; Reeder, 2009a; Van Bavel, Xiao, & Cunningham, 2012). The burden of social warrant puts pressure on moral perceivers to have access to

critical information *content* (causality, intentionality, and so on), but how this information is processed need not be accessible.

Second, the structure depicted in Figure 2 is a conceptual hierarchy of fundamental social-cognitive categories, so their default relationships are indeed conceptual in nature. For example, wondering about *intentionality* makes sense only for events that were brought about by an *agent*, and people care about the agent's *reasons* only for *intentional* behaviors. These relations hold because of how people understand the concepts of *agent*, *intentionality*, and *reasons*. But this conceptual hierarchy translates into a *default processing order* when the information relevant to these concepts must be acquired, probed, or otherwise considered. For example, if the event is underspecified, agency will be probed before intentionality, which will be probed before reasons. (We will offer direct evidence for this prediction later; Guglielmo & Malle, 2013.) But the conceptual relationships also allow for more flexible relations at the process level. For example, at times the perceiver already knows or assumes some "later" information component, or the available information settles multiple concepts at once (e.g., reason information implying intentionality). In such cases the processing order is loosened and the perceiver does not have to plow through each processing step at a time. In a later section (*From Concepts to Process*) we provide more detail on the dynamics of information processing within the overall conceptual structure.

Third, blame judgments should not be pigeonholed as either "rational" or "irrational." They are *systematic* in that they emerge from processing of predictable classes of information that stand in conceptual relations to one another; but they are *defeasible* in that the information processing involved is fallible; the underlying evidence can be unreliable; and, as with all other cognition, arriving at a blame judgment is intertwined with emotion and motivation.

We now discuss each component of the Path Model in detail and review supporting evidence from past research.

## Negative Event Detection

People blame others *for something* (Boyd, 2007). En route to blame, perceivers therefore must first detect an event that violates a perceived norm.<sup>5</sup> This

<sup>5</sup>The theory has no commitments toward drawing a sharp boundary between moral and nonmoral norms because most nonmoral norm violations will trigger the same conceptual and processing mechanism we describe for moral norms, just with less burden of warrant. For example, dressing the wrong way for the opera will generate varying levels of criticism depending on whether it was done intentionally or unintentionally, whether the agent had an option to choose a different attire, and so on.

Type 2 moral judgment may seem to be a trivial constituent of blame, but a number of interesting phenomena occur at this stage.

## Norms

Event detection requires a norm system against which an event is categorized as a violation (Bartels, 2008; Mikhail, 2007; Nichols, 2002). This means that organisms without a norm system are not capable of blaming. The landscape of norms is of course vast and variable and can be partitioned in multiple ways. For example, J. Graham, Haidt, and Nosek (2009) suggested that moral judgments arise in response to distinct domains of violations, including harm, fairness, authority, purity, and ingroup loyalty. Rai and Fiske (2011) asserted that moral norms reflect motives for maintaining and regulating different social relationships. Janoff-Bulman and Carnes (2013) distinguished between proscriptive norms (that identify actions one *should not* perform) and prescriptive norms (that identify actions one *should* perform), which can apply to different targets: self, other, and group. Whatever the most appropriate way of characterizing the norms relevant for moral judgment, detecting an event that violates a norm serves as the critical first step for blame.

## Event Detection Is Simple

Detecting moral events is a much simpler process than making Type 3 judgments such as blame. First, moral event detection does not require theory of mind capacities. Individuals on the autism spectrum can reliably detect norm-violating events (Zalla, Sav, Stopin, Ahade, & Leboyer, 2009) and distinguish different violations from one another, such as interpersonal from property damage (Grant, Boucher, Riggs, & Grayson, 2005), moral from conventional violations (Blair, 1996; Leslie, Mallon, & Dicorcia, 2006), and moral violations from merely disgusting events (Zalla, Barlassina, Buon, & Leboyer, 2011).

Second, even though moral event detection is typically accompanied by evaluative responses ("this is bad"), these evaluations are not necessarily affectively rich, or affective at all (cf. Niedenthal, Rohmann, & Dalle, 2003). Recent work has shown that psychopaths, who do not have emotional responses to others' distress (e.g., Blair, Mitchell, & Blair, 2005), are in fact capable of recognizing and distinguishing moral violations (Blair, 1999; Dolan & Fullam, 2010; Harenski, Harenski, Shane, & Kiehl, 2010),<sup>6</sup> including the popular difference between

<sup>6</sup>Blair (1995) reported that psychopaths were not able to distinguish moral and conventional norm violations, but subsequent studies showed that they were able to (Maxwell & Le Sage, 2009).

“personal” and “impersonal” violations (Cima, Tonnaer, & Hauser, 2010; Koenigs, Kruepke, Zeier, & Newman, 2012). Even though psychopaths do not *care* about norms (Cima et al., 2010; Maxwell & Le Sage, 2009), they do recognize and differentiate norm violations.

Similarly, patients with lesions in their ventromedial prefrontal cortex are characterized as having disturbed emotionality (showing blunted emotional experience, apathy, lack of empathy; Barrash, Tranel, & Anderson, 2000), a condition sometimes dubbed “acquired psychopathy” (Blair & Cipolotti, 2000). But they, too, have no trouble detecting and differentiating norm violations of various kinds, such as moral vs. conventional (Saver & Damasio, 1991), personal versus impersonal (Ciaramelli, Muccioli, Làdavas, & di Pellegrino, 2007; Koenigs et al., 2007; Moretto, Làdavas, Mattioli, & di Pellegrino, 2010), and direct versus indirect harm (B. C. Thomas, Croft, & Tranel, 2011).

Thus, it seems clear that detecting norm violations and recognizing which norm is violated is a simple, nondemanding process for the human mind.

### *Variety of Events*

Norm-violating events come with varying amounts of information. When the event is an *outcome* (e.g., a scratch on one’s car door), very little is revealed, not even whether an agent is involved. When the event is a *behavior*, agent causality is assured and information processing can immediately focus on intentionality. The same is true for “nonbehaviors” such as omissions or intentions; letting someone die or planning to hurt someone are not physical movements, but they imply the involvement of an agent, and the intentionality concept is activated.

Some norm-violating events are so prototypical that subsequent concepts’ values are instantly set and information processing is sped up (Fransson & Ask, 2010). For example, learning that a school shooting occurred leaves no question about agent causality and intentionality, nor would anyone wonder whether the agent’s reasons for acting could justify the action. All the relevant information is available upon detecting the event and appropriate blame can ensue.

Finally, sometimes moral perceivers face compound events, such as when a plan for one outcome goes awry and a different outcome ensues. Such events can combine neutral plans with mildly harmful outcomes or mischievous plans with terrible outcomes, occasionally even vicious plans with harmless outcomes. Moral perceivers are able to assess both the manifest (the norm-violating outcomes) and the representations (e.g., norm-violating intentions), and they systematically integrate the two (Cushman, 2008).

### *The process of event detection*

The mental process of detecting (and often evaluating) a norm-violating event may rely in part on the operation of moral “intuitions” based on “moral grammar rules” (Haidt, 2001; Mikhail, 2007). Some norm violations—direct physical harm to another person, for example—are quickly detected, and perhaps more strongly weighted, with the help of somatic responses (Cushman, Gray, Gaffey, & Mendes, 2012; Damasio, 1994). More generally, people are highly sensitive to negative events. Compared with positive or neutral events, negative events command more attentional resources, are more widely represented in language, and exert a stronger impact on interpersonal behavior (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Ito, Larsen, Smith, & Cacioppo, 1998; Rozin & Royzman, 2001; Taylor, 1991). Once detected, such events can trigger rapid evaluative responses (Luo et al., 2006; Van Berkum, Holleman, Nieuwland, Otten, & Murre, 2009) and activate the moral judgment machinery by flagging the types of norm violations that are worthy of further processing (Mikhail, 2007).

But a rapid negative evaluation that “something bad happened” does not constitute a judgment of blame (Pomerantz, 1978). Blame arises in part from assigning meaning to an event—a fundamental process in social cognition. Finding meaning answers a *why* question, resolving uncertainty by filling a gap in understanding (Hilton, 2007; Malle, 2004). People experience nagging *why* questions for a variety of events, but particularly for negative ones (Malle & Knobe, 1997a; Wong & Weiner, 1981). Thus, detecting a negative event almost inevitably elicits an attempt to find its meaning; and blame requires meaning of a particular kind—one that involves an *agent* who *caused* the negative event.

### *Agent Causality*

For blame to emerge from the detection of a negative event, the perceiver must establish that an *agent* caused the event (Shaver, 1985; Sloman, Fernbach, & Ewing, 2009). Numerous studies have demonstrated the crucial role of agent causality in assigning blame (Cushman, 2008; Lagnado & Channon, 2008) and for social perceivers from age 5 on (Shultz, Wright, & Schleifer, 1986).

The agency concept, emerging early in infancy, relies on features such as self-propelledness and contingent action (Johnson, 2000; Premack, 1990). That is not enough, however, to qualify as a *morally eligible* agent. Such moral eligibility requires that the violated norm *applies to* the agent by virtue of her role or identity (Schlenker et al., 1994) and that the agent is able to understand and remember norms to appropriately modify her behavior through intentional

control (Guglielmo et al., 2009). If such abilities are absent (e.g., in infancy or in certain mental or physical illnesses) blame will either not be assigned or be decisively mitigated, in everyday life as in the law (Alicke, 1990; Monroe, Dillon, & Malle, 2014; Robinson & Darley, 1995, Chapter 5).

In most situations, agent causality will take on a dichotomous Yes/No value. Other situations will call for a graded value: when moral eligibility is partial or uncertain (e.g., a 12-year-old murderer) or when causality is distributed across multiple agents or causal factors (Spellman, 1997). But even just a modest value of agent causality should suffice to activate the next concept in the framework of blame: intentionality. Regardless of how large an agent's causal contribution, the social perceiver will want to know whether that contribution was intentional or unintentional.

### Intentionality

The Path Model postulates that an agent's causal involvement falls into two fundamentally different categories—intentional and unintentional (Heider, 1958; Malle, 1999; Reeder, 2009b; White, 1995). Recognizing a behavior as intentional is a core capacity of human social cognition (Malle, Moses, & Baldwin, 2001). It originates in infants' ability to recognize goal-directed motion (Wellman & Phillips, 2001; Woodward, 1998) and to segment the behavior stream into intention-relevant units (Baldwin, Baird, Saylor, & Clark, 2001). The intentionality concept is refined by children's emerging understanding of desire by age 2 (Meltzoff, 1995; Repacholi & Gopnik, 1997), belief by age 4 (Moses, 1993; Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983), and intention by age 6 (Astington, 2001; Baird & Moses, 2001). This differentiation culminates in an adult concept of intentionality that encompasses five components—desire, belief, intention, skill, and awareness (Malle & Knobe, 1997b). Even though people are highly sensitive to these five components in moral and nonmoral domains (Guglielmo & Malle, 2010a, 2010b; Malle & Knobe, 1997b, 2001), they do not deliberate about the components each time they judge whether a behavior is intentional. Instead, they quickly recognize intentionality in everyday situations (Barrett, Todd, Miller, & Blythe, 2005; Malle & Holbrook, 2012), often relying on perceptual cues (Scholl & Tremoulet, 2000) or scripts (Schank & Abelson, 1977), and, for prototypical stimuli, determine intentionality within a few hundred milliseconds of detecting a behavior (Decety & Cacioppo, 2012).

Intentionality judgments are pivotal to social cognition, regulating attention in interaction (Carpenter, Akhtar, & Tomasello, 1998; Malle & Pearce, 2001), as well as guiding explanations (Malle, 1999) and

predictions of behavior (Malle & Tate, 2006). Equally important is their role in moral judgment, as people consistently blame intentional norm violations more severely than unintentional ones (Darley & Shultz, 1990; Gray & Wegner, 2008; Lagnado & Channon, 2008; Ohtsubo, 2007; Plaks, McNichols, & Fortune, 2009; Young & Saxe, 2009; see Dahourou & Mullet, 1999; Ohtsubo, 2007, for non-Western samples). Children as early as age 5 understand that doing something bad intentionally is worse than doing it unintentionally (Karniol, 1978; Shaw & Sulzer, 1964; Shultz et al., 1986; Surber, 1977), and criminal law systems across the United States, Europe, Islamic cultures, and China incorporate intentionality into their gradations of crime (Badar & Marchuk, 2013).

Consistent with these data and previous theoretical accounts, the Path Model asserts that intentionality amplifies blame. But the Path Model's novel and unique claim is that intentionality judgments *bifurcate the perceiver's information processing* (see Figure 1). Just as people *explain* intentional and unintentional behaviors in conceptually and cognitively distinct ways (Malle, 2004, 2011), so do they search for and respond to distinct information when morally evaluating intentional as opposed to unintentional events, as described next.

### Intentional Path: Reasons

When moral perceivers regard the negative event in question as intentional (the left path in Figure 2), they consider the agent's particular reasons for acting. People infer reasons with ease (Malle & Holbrook, 2012), and they find it painful not to know the reasons for someone's action (Malle, 2004). Children explain intentional actions with reasons from age 3 on (Bartsch & Wellman, 1989), and by age 4 they can tell whether one and the same action is good or bad depending on the agent's reasons (Baird & Astington, 2004).

Considering an agent's reasons is an intrinsic part of the moral perception of intentional actions because these reasons determine the *meaning* of the action (Binder, 2000; Scanlon, 2008)—what the action reveals about the agent's motives, beliefs, and attitudes (Malle, 2004; Stueber, 2009). Taking into account this social-cognitive information not only characterizes blame as a person-directed judgment but facilitates two other major responses to norm violations: behavior regulation (by intervening effectively on what the agent wants, believes, and cares about) and evasive action (by anticipating what the agent will do in the future).

More specifically, reasons influence the moral perceiver's degree of blame because reasons can justify or aggravate the action in question. Justifications



have been treated mostly as the norm violator's attempt to mitigate blame through impression management (Darley, Klosson, & Zanna, 1978; Semin & Manstead, 1983; Shaver, 1985); but equally important is the moral perceiver's consideration of reasons, whether or not the violator offers them in defense.

Which particular reasons reduce blame by justification or increase blame by aggravation depends on such factors as communal and legal norms (Alexander, 2009, Chapter 4; Shaver, 1985), the perceiver's ideology (Tetlock et al., 2007), and the norm violator's status and role (Polman, Pettit, & Wiesenfeld, 2013; Riordan, Marlin, & Kellogg, 1983). Prototypical reasons that aggravate blame for negative actions are asocial, selfish, or vengeful goals (Reeder, Kumar, Hesson-McInnis, & Trafimow, 2002) and goals that predict further norm-violations, such as stealing money to buy drugs (Tetlock et al., 2007). Prototypical reasons that justify an otherwise negative action include desires to serve a greater good (Howe, 1991; Lewis et al., 2012; McGraw, 1987) and beliefs that one is threatened and therefore permitted to harm another in self-defense (Finkel, Maloney, Valbuena, & Groscup, 1995; Robinson & Darley, 1995). Because it takes time to learn the many shades of justifying and aggravating reasons, children master the justification component of blame only gradually between the ages of 5 and 9 (Fincham, 1982), later than other constituents of blame.

### **Unintentional Path: Obligation and Capacity to Prevent**

When moral perceivers regard a norm-violating event as unintentional (the right path in Figure 2), they process a complex array of information about what should and could have happened, which is distinct from considerations of what caused the event in the first place (Mandel & Lehman, 1996). They consider to what extent the agent had an *obligation* to prevent the negative event (e.g., due to role, relationship, or context) and to what extent the agent had the *capacity* to prevent the negative event (both the cognitive capacity to foresee the event and the physical capacity to actually prevent it). According to the Path Model, only when moral perceivers explicitly ascribe or implicitly assume an agent's obligation and capacity to prevent the event will they blame the agent for the unintentional norm violation.

### **Evidence for the Impact of Obligation**

Most studies of moral judgment hold obligation constant, typically presenting stories in which the agent unquestionably had an obligation to prevent the negative event in question. Consequently, there is sparse direct evidence for the impact of obligation on

blame judgments. When obligations have been empirically examined, however, they have exerted considerable influence. Hamilton (1986) reported that people in higher positions of a social hierarchy are subject to stronger obligations for preventing negative outcomes and are blamed more for those outcomes when they occur. Similar effects of role position were found in organizational contexts when causality was ambiguous (Gibson & Schroeder, 2003) and even in cases of vicarious responsibility (Shultz, Jaggi, & Schleifer, 1987).

### **Evidence for the Impact of Capacity**

The impact of the cognitive capacity to prevent (often labeled *foreseeability*) has been demonstrated in adults as well as children from age 4 on (e.g., Nelson-Le Gall, 1985; Shaw & Sulzer, 1964) and is the basis for the legal concept of negligence. Agents who cause a norm-violating event that they foresaw (or could have foreseen) receive more blame than agents who cause a norm-violating event that they did not and could not foresee (holding physical capacity constant). In addition, Weiner (1995) reviewed numerous studies in which the agent's physical capacity to control an unintentional outcome was a strong predictor of blame. For example, if a person's obesity is caused by an uncontrollable medical condition, people don't consider the person blameworthy for being obese. If, however, a change in diet promises to counteract the person's obesity (even in the presence of the medical condition), the person may be blamed for failing to pursue this course. Critical for the notion of capacity, therefore, is not only which particular factors are seen to have caused the negative event but which alternative options were reasonably available to prevent the event. Indeed, in Creyer and Gürhan (1997), a driver was blamed more for a freak accident when a counterfactual preventive action was made salient (putting on seat belts), and Catellani, Alberici, and Milesi (2004) showed that a perceiver's focus on alternative actions that a rape victim could have taken predicted the perceiver's judgments of preventability and, in turn, blame (for parallel effects on self-blame, see Davis, Lehman, Silver, Wortman, & Ellard, 1996). Similarly, victims of sexual assault or severe accidents (Davis et al., 1996; Janoff-Bulman, 1979; Janoff-Bulman & Wortman, 1977) often blame themselves because they believe they could have prevented the negative outcome (A. K. Miller, Handley, Markman, & Miller, 2010).

### **Relationship Between Obligation and Capacity**

Typically less information is needed to determine obligation (e.g., the agent's role) than to determine

capacity (e.g., the agent's knowledge, skills, tools, opportunities). It would therefore be inefficient for a cognitive system to first assess whether the agent could have prevented the negative event only to realize that the agent had no obligation to prevent it. Moreover, knowledge of obligations is often available as part of the event representation. For example, when a pedestrian is killed in traffic, perceivers immediately know that drivers have an obligation to prevent such events. Considerations of capacity, assuming unintentionality, would then follow. However, sometimes capacity information can strengthen obligation—such as when a person's knowledge about risks creates an obligation to take special care in preventing them—and if the person did not take such precautions, counterfactual thinking (he should have and could have ...) increase blame (Gilbert, Tenney, Holland, & Spellman, 2013).

### Comprehensive Evidence

The research cited so far has provided evidence for the role of specific components of the Path Model of Blame in isolation, but the complete model has not been tested as a whole. A few studies have tested subsections of the model. Boon and Sulsky (1997) showed that when people assess hypothetical breaches of trust in their romantic relationships, blame judgments are acutely sensitive to variations in intentionality and preventability. Participants in Quigley and Tedeschi (1996) recalled a specific instance in which someone had harmed them, and structural equation modeling showed that ratings of harm severity, intentionality, and (lack of) justification predicted blame. Mikula (2003) proposed an "attribution of blame model" of injustice judgments and showed across five studies that judgments of injustice/blame were guided by perceptions of causality, intentionality, and justification. Finally, Jones and Kelly (2010) showed that deleterious effects of being excluded from social information follow the same principles as blame does: Information exclusion was most negative when it appeared intentional; it could be mitigated by justifying reasons; and when the exclusion was unintentional, it was negative only when perceived as preventable.

Beyond this evidence for partial configurations, the first comprehensive tests of the Path Model have been conducted recently in our own lab, and we summarize them next.

### Recent Tests of the Model

#### *Information Acquisition*

Perceivers often lack complete information about negative events and must actively search for

additional information before arriving at a blame judgment. Because of its hierarchical structure the Path Model predicts a default order in which moral perceivers seek out information or prioritize the consideration of different types of information. It holds that upon detecting a negative event, perceivers will first seek information about causality, then (if the event was agent-caused) about intentionality, then (if the event was intentional) about either reasons or (if the event was unintentional) about preventability.

We examined these predictions in two complementary experimental paradigms (Guglielmo, 2012; Guglielmo & Malle, 2014). In both, participants read about a variety of norm-violating events and had opportunities to acquire additional information in order to determine who or what is to blame for the event. In the "information search" paradigm, they were allowed to ask questions about whatever they wished to know (without any guidance as to the kinds of information they might request), and the questions were content coded into theoretically meaningful categories. In the "information offer" paradigm, participants received counterbalanced offers for particular types of information (viz., the critical concepts of the Path Model) and indicated, for each offer, whether they wanted to receive that type of information.

The results of both paradigms supported the Path Model. In the information search paradigm, people asked questions about the relevant types of information in the predicted order. When learning about negative events, people primarily asked questions about agent causality; when learning about agent-caused events, they primarily asked questions about intentionality; and when learning about intentional actions, they primarily asked questions about reasons. Unintentional negative events frequently elicited preventability questions, though they also elicited questions clarifying background details of the event or the potential causal involvement of other individuals.

In the information offer paradigm, participants were fastest and most likely to accept the predicted types of information. For example, upon discovering a negative event, they were most inclined to accept causality information; upon discovering an agent-caused negative event, they were most inclined to accept intentionality information. Moreover, these same patterns emerged even when participants had minimal time (2,000 ms) to accept or reject information, suggesting that the processing outlined by the Path Model can be either deliberative or intuitive.

#### *Information Updating*

The Path Model's hierarchical structure makes unique predictions about the assimilation of new information that expands or contradicts initially

acquired information. Intentionality bifurcates information processing into two distinct paths, each targeting specific informational requirements for blame. On the intentional path, moral perceivers selectively consider reason information; on the unintentional path, moral perceivers selectively consider preventability information. If, during this selective processing, opposing information about intentionality arises, the system must “step back” to the bifurcation point, update the intentionality judgment, and consider information on the other path before the blame judgment is made. Such mental “path switching” will come with processing costs.

We tested this hypothesis by assessing the speed with which people updated their moral judgments for path-switching (compared with path-maintaining) scenarios, presented as either written or auditory stimuli (Monroe, 2012; Monroe & Malle, 2014). Participants received information about a moral transgression (e.g., “Eric broke Monica’s arm,” which most people assume to be unintentional) and made an initial blame judgment. Then participants received new information, which was either path-switching (in the aforementioned case, reason information) or path-maintaining (preventability information). Finally, participants were allowed to update, if desired, their initial blame judgment. As predicted by the Path Model, both student and community members were indeed slower at updating blame in the path-switching scenarios than in the path-maintaining scenarios. Moreover, this effect was not due to a general expectancy violation in the path switching scenarios. A follow-up study showed that people were still slower at updating blame in path-switching scenarios, even when those scenarios were far more common than path-maintaining scenarios.

### From Concepts to Process: The Dynamics of Information Processing

The just reported results illustrate that patterns of information seeking and information updating are highly systematic and conform well to the Path Model’s predictions. Building on these results, we now introduce a second layer of the Path Model, which can be independently falsified. It concerns the specific information processes that occur *at each conceptual node* in the larger conceptual structure (e.g., *agent causality, intentionality*).

#### Information Processing at Each Conceptual Node.

Up to three elements of information processing occur at each conceptual node:

Concept activation → Information acquisition →  
Value setting (CIV).

In brief, once a concept is *activated* the system acquires concept-specific *information*, which is used to set the concept’s *value* (cf. Gawronski & Bodenhausen, 2006). Thus, here too, the Path Model postulates a conceptual hierarchy that translates into a processing order *to the extent* that processing occurs (more on this qualification shortly).

*Information acquisition* can consist in active information search (e.g., probing an agent’s causal involvement), knowledge retrieval (e.g., recalling the agent’s role and obligations), perception (e.g., reading the word “intentionally” or seeing a certain movement configuration), inference (e.g., what the reasons might be for the focal action), or simulation (e.g., what the agent could have done to prevent the event). The Path Model of Blame does not constrain which of these modes of acquisition will lead to the desired information. We have seen in Guglielmo and Malle’s (2013) findings that, at the level of active information search, the ordering postulated by the Path Model is well supported. Additional studies will be needed to examine this ordering at more implicit levels, such as by way of eye-tracking data.

*Value setting* can be thought of as exceeding a subjective probability threshold that the relevant criterion is met, such as  $p(\text{agent caused event})$  or  $p(\text{reasons were justified})$ . As soon as the value of one concept is set, it activates the next concept in the hierarchy. For example, once it is established that an agent caused the event in question (agent causality value is set), the intentionality concept is activated and relevant information acquisition begins until threshold—for example, for  $p(\text{behavior was intentional})$ —is reached.

#### Parsimony

The information acquisition and resulting value setting processes will not always occur for each and every concept one at a time; we assume that the system processes information parsimoniously (Fiske & Taylor, 1984), leading to at least four kinds of “shortcuts.”

1. *Hierarchy*. For any given concept, if information is already available, the concept’s value is set, and processing can focus on the as yet uncertain other concepts. Because of the hierarchical conceptual structure of blame, only concepts further down from the preactivated concept need to be considered.
2. *Event-implied information*. Parsimony can arise already at event detection, when information relevant for subsequent concepts is mentioned, observed, implied, or assumed. For example, when we see a teenager bump into someone on the sidewalk, briefly hold a wallet, and dash off, the pickpocketing script will likely be activated

(Schank & Abelson, 1977), setting the intentionality parameter to *Yes* and justification by reasons to *No*. Hearing someone say that “he forgot his wife’s birthday” implies (by verb choice) a lack of intentionality and (by way of role term) an obligation value of *Yes*, since spouses, in this culture, are expected to remember each other’s birthdays. Finally, observing some norm-violating events can activate schemas that don’t directly set values but narrow the perceiver’s search for relevant information. If a dog bites a child in the park, one may quickly search for the dog owner as a potential causal agent with an obligation to prevent this kind of event.

3. *Multiple-concept information*. Some pieces of acquired information can set the values for multiple concepts. Seeing that a person has a badly injured finger and learning that this occurred because “somebody tried to steal her diamond ring” implies a causal agent, intentionality, and a clearly unjustified reason. In this case, there is no need to acquire information about each of these concepts separately—the event provides them all at once.
4. *Preset values*. An intriguing shortcut in the blame process occurs when values are “preset” by activated knowledge structures. Preset values may be associated with specific agents (e.g., Monisha tends to be reckless), roles (e.g., dentists have an obligation to prevent patients’ pain), or group memberships (e.g., the rival always intentionally harms us). Concept values can also be preset in certain perceivers. Children, for example, assume that positive outcomes tend to be intentional (Jones & Thomson, 2001), and people who see rape as a sexual act rather than an act of violence assign greater partial causality to the victim (McCaul, Veltum, Boyechko, & Crawford, 1990).

In all four types of shortcuts, people show rapid moral judgments because they do not have to go through a multistep process of acquiring the relevant information. This may be the information-processing basis for what has been called “intuitive” moral judgments. For example, empirical tests of Haidt’s (2001) model typically use narratives in which causal agency, intentionality, and justifications are made patently obvious (J. Graham et al., 2009; Haidt & Hersch, 2001; Wheatley & Haidt, 2005). In such cases, the perceiver has little computational work to do between recognizing the norm violation and forming a moral judgment (even a Type 3 judgment), because all concept values are already provided in the stimulus. We should not conclude from such cases, however, that people always “intuit”

moral judgments directly, without processing the critical information identified in the Path Model. Many everyday events are sparse and require further processing, in which case people seek to acquire information about causal agency, intentionality, justified reasons, and the like.

Spelling out the CIV dynamics also allows for more precise analyses of how affect and emotion are involved in the emergence of blame, and we will return to this issue.

### Part 3: Alternative Theoretical Approaches

We now compare the Path Model with past and present theories of blame and well-known claims about blame.

#### Why Omit the Responsibility Concept?

Many previous models of moral judgment assigned a central role to the concept of responsibility (Fincham & Jaspars, 1980; Schlenker et al., 1994; Semin & Manstead, 1983; Shaver, 1985; Weiner, 1995). Why not our model? We omit responsibility because it is a hopelessly equivocal concept (Feinberg, 1970; Fincham & Jaspars, 1980; Hamilton & Sanders, 1981; Hart, 1968; Sousa, 2009). It collapses distinct phenomena under a single label and is often confounded with other phenomena. A recent study shows at least four constructs that are subsumed under or co-measured with responsibility: wrongfulness, causality, foreknowledge, and intentionality (Gailey & Falk, 2008). In addition, the term *responsibility* has been used to refer to an agent’s obligation (Hamilton, 1986), eligibility for moral judgment (Oshana, 2001), intentionality and justification (Fincham & Bradbury, 1992), and simply blame. For example, Shaw and Sulzer (1964) suggested that “When one person attributes responsibility for an event to another individual, he blames that person if the outcome is negative” (p. 39). Likewise, Shultz, Schleifer, and Altman (1981) told their participants that “*moral responsibility* refers to the extent to which the protagonist is *worthy of blame*” (p. 242). Conversely, Fincham and Shultz (1981) told their participants that “blame concerns the extent to which someone should be held morally responsible” (p. 115), and Quigley and Tedeschi (1996) measured the construct of blame by asking participants about responsibility. But responsibility measures are less sensitive than blame measures to manipulations of various determinants of moral judgment, such as intention, foreseeability, and justification (e.g., Critchlow, 1985; McGraw, 1987). This is most obvious for cases in which an agent’s

intentional action violates a norm but is either justified or not justified by a good reason. In both cases the agent is “responsible” for the action but only in the second case does he deserve blame (Heider, 1958; Shaw & Sulzer, 1964).

For all these reasons we have omitted the term *responsibility* from our model and included instead the more precise concepts with which it has been confounded: causality, intentionality, and obligation.

### Cushman’s (2008) Model of Wrongness and Blame

A recent model of moral judgment offers an important distinction between two kinds of moral judgments: wrongness and blame. Cushman (2008) stated that people’s judgments about the *wrongness* of an agent’s behavior are driven by assessments of the agent’s mental states—namely, the agent’s beliefs and desires. Thus, people judge a behavior to be especially wrong when the agent believes his behavior will bring about a negative outcome and wants this outcome to occur (regardless of whether the outcome actually occurs). Judgments of *blame*, however, also take into account the actual consequences of the agent’s behavior—whether a negative outcome in fact occurred. In this way, an agent receives more blame for a behavior that happens to have bad consequences than for one that does not, holding constant the agent’s mental states (Mazzocco, Alicke, & Davis, 2004; Robbenolt, 2000). Still, mental state judgments remain critical for assignments of blame, holding constant the consequences: An agent who lacks either the relevant belief or desire and thus unintentionally causes a negative outcome will be blamed much less than an agent who has the relevant belief and desire and intentionally caused that outcome (Cushman, 2008).

Cushman’s model and our Path Model share important features, but they do differ in several respects. First, Cushman did not specify how people are blamed for unintentional behaviors. His model predicts only that in the absence of intention, blame will be low. But blame is not uniformly low in such cases; considerations of the agent’s obligations and capacities are critical in blaming unintentional behavior. Second, Cushman did not distinguish between mental states that function as reasons for acting intentionally and mental states that represent the cognitive capacity to prevent negative outcomes (e.g., believing that one’s action may have a negative side effect). Finally, Cushman’s model does not distinguish between justified and unjustified reasons, both of which bring about an undesirable intentional action but only the latter of which leads to blame.

More generally, however, Cushman’s model raises important questions about the relationship between

wrongness and blame that research has not yet addressed. For one thing, is wrongness a judgment *sui generis* or is it equivalent to a blame judgment of norm-violating *actions*? (Unintentional events are unlikely to be called “wrong.”) Moreover, are norm-violating actions that are done for justified reasons (e.g., killing out of self-defense) considered “wrong”? Examining this question will reveal whether people process detailed reason content when assessing wrongness or focus on the *type* of action (e.g., lying is always wrong, even though lying to protect the other person’s feeling does not deserve blame); and it might reveal whether justified norm-violating actions, though “officially” blameless, might still leave the moral perceiver with a twinge of negative evaluation. People may not escape the impression that the agent performed a wrong *type* of action, even if for the right reasons.

### Dual-Process Model of Permissibility

Greene (2007, 2009) suggested that people have immediate aversive emotional reactions to so-called “personal” norm violations (e.g., those involving direct physical harm) and are inclined to judge such violations as morally impermissible. People also often engage in deliberate conscious reasoning, which may temper their initial negative emotional reactions to those violations. These two processes—one automatic and emotional, the other deliberative and reason-based—normally unfold in parallel, such that people’s ultimate moral judgments are guided by whichever processing stream wins out over the other. In particular, Greene suggested that emotional processing tends to favor “deontological” moral judgments (i.e., that a given action is wrong, regardless of its consequences), whereas deliberative processing tends to favor “consequentialist” moral judgments (i.e., that a given action is wrong in proportion to its negative consequences).

Greene’s model is supported by evidence demonstrating that heightened activation in brain regions believed to subserve emotions predicts deontological judgments, whereas heightened activation in brain regions believed to subserve reasoning predicts consequentialist judgments (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Moreover, ventromedial prefrontal cortex patients—who have diminished emotional reactions—make more utilitarian judgments (Koenigs et al., 2007), and so do healthy participants who have experienced a positive mood induction (Valdesolo & DeSteno, 2006).

Recent studies suggest a more complex picture. One study found that participants’ emotions did not predict how participants resolved a moral dilemma, but cost–benefit calculations for various alternative

action paths did (Royzman, Goodwin, & Leeman, 2011). Another study examined how induced stress would affect people in resolving moral dilemmas, predicting that higher stress leads to overweighting the emotion-favored action path (Youssef et al., 2012). But stress (measured with cortisol levels) led to only marginal increases in rejecting emotion-inducing “personal” violations (79–86%; derived from graphed means) and to identical increases in rejecting *impersonal* violations (39–44%), which are hypothesized to involve little emotional processing. Moretto et al. (2010) found that affective reactions (measured by skin conductance) were present only when people decided to *accept* personal violations (for utilitarian reasons of saving several lives), contradicting the hypothesis that quick, automatic affect guides people to *reject* those violations (Greene, 2007). Participants in Moretto et al.’s study deliberated longer when they endorsed the utilitarian option (see also Greene et al., 2004), but this seems to reflect the act of weighing the conflicting options (Baron, Gürçay, Moore, & Starcke, 2012). In fact, (Koop, 2013), using a mouse-tracking methodology, found no indication that deontological responses were faster than utilitarian ones. Affect seems to be part and parcel of reasoning about moral events, not a shortcut that somehow bypasses reasoning.

Even with adjustments to accommodate these findings, Greene’s dual-process model does not account for judgments of blame. First, the model is tailored to a particular class of events—moral dilemmas that create a conflict between fast intuitive reactions and controlled deliberations; how people make moral judgments for everyday norm violations is not specified. Second, the model is tailored to one kind of moral judgment—assessments of (im)permissibility, which are Type 2 judgments in our classification, measuring norm violations at the event detection stage of blame formation. Third, the deontological/consequentialist distinction, central to Greene’s model, does not seem to make a difference for how blame comes about. When people judge agents as blameworthy, they are not doing so in a deontological or consequentialist manner. A perceiver may identify a behavior (e.g., pushing) as violating a deontological norm (“pushing is wrong”) or a consequentialist standard (“this instance of pushing has no benefits”); either way, for people to assign actual blame they still need to consider information about agent causality, intentionality, preventability, and so on.

Which of the two demarcated processing paths—affect or deliberation—takes in such blame-relevant information? It seems uncontroversial to assume that the deliberation path can do so. But Greene, Morelli, Lowenberg, Nystrom, and Cohen (2008) also consider the possibility that the affective-intuitive processing path is sensitive to intentionality, reasons, and

similar considerations. In fact, Greene et al. (2009) showed that a presumed trigger of affective processes (i.e., personal force) had an impact on permissibility judgments only for intentional, not for unintentional, behaviors. Similarly, Decety, Michalska, and Kinzler (2012) found that activation in the amygdala (often described as subserving emotion processing; Adolphs, 1999) was highly sensitive to the intentionality of observed immoral behaviors. Both of these possibilities—that blame-relevant information gets processed by controlled deliberation or by affective intuition—are accommodated within the Path Model of Blame, for which the *kind* of information is critical, not the mode by which it is processed.

We now turn to an apparent challenge to our model that doesn’t come from one particular theory but from the widespread claim that moral judgment is subject to motivational biases—in particular, that people have a desire to blame, which distorts their default information processing. We begin with the classic hypothesis of outcome bias.

### Motivated Blame 1: Outcome Bias

Early research on responsibility attribution examined motivated moral judgments for accidents and misfortunes (Shaver, 1970; Walster, 1966; for reviews, see Burger, 1981; Robbennolt, 2000). The initial hypothesis was that severe misfortunes (e.g., a person being assaulted on the street) threaten an observer’s sense of control. To restore this sense of control the observer tends to see the misfortune as more preventable and therefore blames the *victim* more for severe outcomes. Increasingly, the hypothesis has turned into a general claim of *outcome bias*—that assessments of blame are distorted by the severity of the outcome (Alicke, 2000; Mazzocco et al., 2004).

This hypothesis, however, has suffered many setbacks. Early studies that showed the impact of outcome severity on responsibility (or blame) judgments were difficult to replicate. More and more moderator variables had to be added to the hypothesis, and the body of research was highly inconsistent (Fishbein & Ajzen, 1973; Shaver, 1970). A meta-analysis of the hypothesis showed that the average correlation between outcome severity and moral judgment was  $r = .08$  for responsibility and  $r = .17$  for blame judgments (Robbennolt, 2000).

There is, of course, an impact of outcome or consequences on blame (e.g., Cushman, 2008). A driver bumping a pedestrian and a driver killing a pedestrian violate different and differentially stringent norms. The puzzle of “moral luck” arises when one imagines that the two drivers had exactly the same mental states, behaved exactly the same way, but differed in the severity of the outcome (Athanasoulis, 2005). Outside of thought experiments, however, how

realistic is it to assume exactly the same mental states? It seems reasonable to infer that more extreme outcomes are usually caused by greater negligence (e.g., less attention, weaker preventive efforts) or, in the case of intentional action, by more extreme motives and committed plans. Outcome bias studies often assumed to hold constant such mental states rather than actually measuring them as potential mediators of the outcome–blame relationship. In one early exception (Fincham, 1982), outcome severity in fact predicted mental state inferences (about the agent’s desire to damage), and these inferences predicted blame judgments. Likewise, in studies that found notable outcome effects on blame (Howe, 1991; Howe & Loftus, 1992), mental state manipulations explained 6 times more variance in people’s blame ratings than did outcome manipulations. More recent studies show the same pattern (Darley, Solan, Kugler, & Sanders, 2010; Young, Nichols, & Saxe, 2010). Thus, the hypothesis of a general undue impact of outcome on blame—because people suspend information processing—is not well supported.

Still, some authors suggest that people’s mental state inferences themselves may be biased—distorting “the facts” in service of a desire to blame (Ames & Fiske, 2013; Mazzocco et al., 2004). Indeed, several recent models have proposed that blame (or something close to it) precedes and generates biased assessments of causality, mental states, and harm. Such “blame-early” models propose that “judgments that an individual is “bad” or “good” often come prior to rather than as a product of more fine-grained judgments of intentionality, controllability, and causality” (Ditto, Pizarro, & Tannenbaum, 2009, p. 316).

## Motivated Blame 2: Blame-Early Models

### *Culpable Control*

The most explicit model of blame-early processing comes from a sustained research program by Alicke and colleagues (Alicke, 1992, 2000, 2008; Alicke, Rose, & Bloom, 2011; Alicke & Zell, 2009). Alicke described two major elements of judgments of blame: evaluations (of the behavior, the actor, and the outcome) and assessments of three “linkages”—how the actor’s mind controlled the actor’s behavior, how the actor’s behavior controlled the outcome, and to what extent the actor’s mind did and should have anticipated the outcome. These three linkages are also referred to as processing of “evidential information.”

Although the terminology is different, Alicke’s Culpable Control Model (CCM) can be mapped onto the Path Model (PM) of Blame, with the latter making some distinctions that the CCM does not make:

- behavior–outcome link  $\sim$  agent causality
- mind–behavior link  $\sim$  combines intentionality and reasons
- mind–outcome link  $\sim$  combines prevention obligation, capacity, and attempts.

Further, both models grant that the moral perceiver performs complex information processing en route to a final blame judgment. Yet there are significant divergences between the PM and the CCM: (a) in whether information processing occurs hierarchically (PM) or simultaneously (CCM), (b) whether intentionality bifurcates information processing (PM) or merely provides evidence (CCM), (c) whether evidential information processing comes early (PM) or late (CCM), and (d) whether information processing is generally evidence based (PM) or generally distorted by extraevidential information and a desire to blame (CCM). We have provided empirical support favoring the PM on the first two points (see the *Recent Tests of the Model* section), so we focus here on the last two points, which put the CC model’s motivated reasoning proposal in relief.

As depicted in Figure 3, early spontaneous evaluations of (evidential and extraevidential) information, such as the actor’s character or the degree of harm, are said to trigger a desire to blame, which in turn distorts evidential information processing (i.e., of causality, mental states) to arrive at the desired level of blame (Alicke et al., 2011, p. 675). We offer two theoretical comments first, then we turn to the evidence.

The explanatory force of the “desire to blame” in the CCM is not entirely clear. In some sense every action, including blaming, has an underlying desire. And even if people were found to process information in the most normative and accurate ways, they would still have such a desire to blame. However, Alicke assumed that the desire to blame seeks *exaggerated* blame (see also Ames & Fiske, 2013; Tetlock et al., 2007). To say that blame is *exaggerated* requires a normative model of blame.

Even though Alicke rejected normative models of blame (e.g., Alicke et al., 2011, p. 671), he adopted a

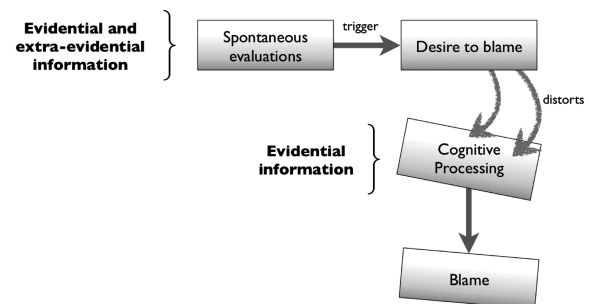


Figure 3. Our depiction of the Culpable Control model of blame. (Color figure available online.)

normative distinction between “evidential” factors (e.g., behavior, causal contribution, intentionality, motives), which *should* influence people’s blame, and “extraevidential” factors, which *should not* influence blame. He identified “philosophers, legal theorists and psychologists” (Alicke, 2008, p. 179) as the originators and arbiters of this normative distinction. Unfortunately, those arbiters often do not agree with one another. For example, Alicke suggested that taking into account the different consequences of two otherwise identical actions is an “outcome bias.” For a utilitarian, however, consequences are the only acceptable basis for ethical judgment. Moreover, among other sources of information, which of these are uncontroversially extraevidential? A history of child abuse? Race? Looks? Past record? Without a consensual and reliable criterion for what is evidential and what is extraevidential, it may be most fruitful to examine the precise psychological processes that lead from event perception to a judgment of blame (N. H. Anderson, 1991; Pepitone, 1975), without the evaluative language of bias and distortion. However, because of the prominence of this language in contemporary psychology we also assess to what extent the current empirical evidence can support charges of distortion.

**Extra-evidential outcome information.** One line of evidence for the impact of a desire to blame on information processing stems from the hypothesis of outcome bias. We have mentioned that outcome effects are small (Robbennolt, 2000), typically evidential, and often readily explained by causal and mental state inferences mediating the outcome–blame relationship. Alicke and collaborators, however, have offered provocative studies to suggest that many mental state inferences that seem to mediate the outcome–blame relationship are in fact post hoc justifications of initial negative evaluations (Alicke, 1992; Mazzocco et al., 2004).

In one set of studies (Alicke & Davis, 1989; Mazzocco & Alicke, 2005), participants read about a homeowner who heard noises in the house, noticed a man going through his daughter’s dresser; and, when the presumed intruder turned around, shot and killed the man. Participants who learned that the killed man was a burglar with a long criminal record blamed the homeowner less than those who learned that the man was the daughter’s boyfriend (who was picking up some clothes for her). This effect of the outcome manipulation on blame was almost entirely mediated by ascriptions of negligence—inferences that the homeowner should have taken preventive steps but did not. Were those inferences of negligence fabricated to justify a desire to blame or were they based on evidence? Enzle and Hawkins (1992) showed, using very similar vignettes, that people

spontaneously make such inferences from both implicit and explicit evidence for negligence, which then determine degrees of blame. But even if one favors a “bias” interpretation, the bias is in the wrong direction. In studies that contained a control group (offering no information about victim identity), the very bad condition typically showed no significant increase in blame relative to the control group (contradicting a desire to blame account), whereas the less bad condition showed a significant decrease in blame relative to control (Alicke & Davis, 1989; Mazzocco et al., 2004).

Furthermore, many outcome bias studies contain a significant confound. The agent who causes the less bad outcome typically has a true belief (e.g., the homeowner correctly believing that a burglar is in the house), whereas the agent who causes the very bad outcome has a false belief (Young et al., 2010). When perceivers learn this fact—that reality turned out to be very different from what the agent believed—they may wonder whether the original belief was reasonable and justified, and if it wasn’t, this would increase blame via the cognitive capacity component (i.e., the agent could have gathered information more carefully or judged the situation more prudently). This is just what Young et al. (2010) showed. People inferred that agents with false beliefs were less justified in their assumptions than agents with true beliefs, irrespective of outcome; for neutral outcomes, false beliefs led to significantly more blame than true beliefs. Further, in cases directly comparable to Alicke’s, bad outcomes and neutral outcomes led to indistinguishable degrees of blame when holding constant false beliefs. Thus, the typical outcome bias effect appears to be driven not by the occurrence of bad outcomes but by the fact that such outcomes reliably indicate false beliefs and therefore elicit considerations of prevention capacity.

In sum, theoretical examination and empirical examination of outcome bias studies provide little support for blatant motivated reasoning in blame judgments. Instead, findings are consistent with two elements of the Path Model of Blame: Outcome information can have an impact because it specifies what the norm-violating event really is and because it reveals something about the agent’s mental states, which are then the primary determinants of blame.

**Extra-evidential agent information.** Besides consequences, the norm violator’s character and ancillary motives are often portrayed as extraevidential and as biasing blame (Alicke, 2000; Landy & Aronson, 1969). In one frequently cited study, Alicke (1992) found that a character who was speeding *in order to hide cocaine* was judged more causally responsible for an ensuing car accident than was a character who was speeding *in order to hide a gift for*



*his parents*. In this case, the outcome is held constant but the agent's mental states (his reasons for speeding) are varied. Alicke (1992) argued that those mental states are irrelevant to the resulting degree of blame for the accident, so using them constitutes bias. However, in real life an agent's goals (and inferred character) may provide preventability information: for example, that the drug-hiding agent was driving faster, was more inattentive, and more careless than the gift-hiding agent, warranting greater causality and blame judgments. We do not know whether participants made such inferences, because they were not measured in the studies.

Another study (Nadler & McDonnell, 2012, Study 2) described an explosion in Sam Norton's garden shed, which killed a neighborhood teenager. Norton's shed posed a significant risk because it was full of oxygen tanks, so the question was how blameworthy Norton was for this accident, as a function of three possible pieces of agent information. Norton had stored the oxygen in the shed for a neutral reason (he is a businessman providing in-home delivery of healthcare equipment), a bad reason (he is a football coach illegally administering oxygen to his players), or a laudable reason (he is a father caring for his daughter who has a respiratory disease). Compared with the neutral condition, participants in the bad-reason condition judged Norton more blameworthy and those in the good-reason condition less blameworthy. This polarizing effect is inconsistent with the specific claim of a "desire to blame." It appears that people made inferences from the agent's reasons whether good or bad. In fact, Nadler and McDonnell (2011, p. 284) pointed out that in the law such information must be taken into account when judging criminal liability (Model Penal Code §§ 2.02(2)(c), (d); American Law Institute, 1985): "When an individual disregards a substantial risk and the nature and purpose of that disregard is not legitimate, that individual may be criminally liable." This undermines the charge of bias in people's moral judgments: If the actual legal prescription is to integrate relevant causal-mental information into the overall judgment, then people do what they are expected to do—or rather, the law has codified ordinary information-processing regularities.

A stringent test of motivated moral judgment would need to separate the extraevidential information source from the norm violation in such a way that no *diagnostic* information (relevant to an interpretation of the norm violation) can be inferred from the extraevidential information. Such a separation might succeed if we could find a direct effect on blame simply because the agent is dislikable. Alicke and Zell (2009) compared a likeable to a dislikeable agent and introduced the respective personalities through facts that were causally separated from the blameworthy event. Personality impressions had the

predicted effect on blame, such that dislikable agents received more blame for accidentally punching a woman (Study 1) or accidentally hitting a bicyclist with his car (Study 2).

However, whether these efforts to separate personality information from the norm-violating event were successful is open to debate. For example, in the critical scene of Study 1, the agent mistook an act of sympathy between a brother and a sister for an act of aggression and, against the woman's assurance that everything was fine, the agent got into a fight with the man, eventually punching the woman accidentally in the face. What information do participants have available to interpret the scene? The dislikable person was, earlier in the day, rude to a policeman, pushy and mean to a friend, drank a few beers, made up an excuse to get out of work the next day; the likable person was polite, contrite over a mistake, helped a friend, and volunteered at a homeless shelter. Of these two agents, who is more likely to make an honest perceptual mistake in the confrontation scene? Whose prosocial motives are in doubt? A convincing study needs to measure participants' inferences regarding these questions and include them as potential mediators.

Nadler (2012) went some way toward such a comprehensive study, manipulating and measuring character and recklessness as well as inferred causal-mental variables. Although concerns can be raised about the lack of a control group and about diagnostic information in the character description, we want to emphasize an intriguing finding: When character was manipulated between subjects, it had the predicted effect on blame, but when it was manipulated within subjects, the effect disappeared entirely. The author interprets this result as suggesting that character influences blame unconsciously (and when it is made conscious, people correct for it). But another view is that people can better distinguish between causally relevant and irrelevant factors in a within-subject design. When two agents with very different character cause identical outcomes, then character is unlikely to be the relevant cause, whereas constant factors (such as recklessness) are likely causes. When people have no such opportunities of comparison (in a between-subjects design), they integrate any and all information given to them, including clues about potentially relevant general dispositions (Tannenbaum, Uhlmann, & Diermeier, 2011), to interpret the causal-mental facts of a naturally ambiguous situation. And that will be of particular importance when judging strangers about whose beliefs and desires the moral perceiver has no background knowledge (Bloom, 2011).

In fact, to properly assess the significance of character information we need to keep in mind that for moral judgments in everyday life (and indeed, in small-group living in our evolutionary past), such

character information is normally available when people evaluate causality, intentionality, and reasons. Nobody would want ordinary perceivers to ignore such base rates about a colleague, friend, spouse, or child. So when people try to draw inferences from the information offered in experiments, they seek out the kind of information that normally helps them strengthen their judgments.

As a result, vignette studies that try to demonstrate the undue effect of extraevidential information face a nearly insurmountable challenge: Because people have to make judgments about ambiguous material, they are inferentially hyperactive and will inspect any information they receive for signs of what they want to know: the agent's causal role, mental states, obligations, preventive actions. Experiments without a ground truth will therefore have a difficult time making the normative distinction between justified and unjustified ("motivated") inferences. One approach for future research might be to manipulate extraevidential information that, according to a desire-to-blame account, should influence all components of blame (e.g., bad character influencing perceived causality, intentionality, reasons, etc.) but that, according to a diagnostic inference account, should influence specific components of blame (e.g., physical strength influencing inferred causality; a caring character influencing inferred motives). A hint of component-specific processing lies in Nadler and McDonnell's (2011) and Nadler's (2012) studies, in which causality inferences were not responsive to character manipulations but mental inferences were responsive.

From the perspective of the Path Model of Blame, people seriously consider any available information (including character) that reveals something about the blame-relevant components of causality, intentionality, reasons, and preventability. Positive evidence for the systematic way in which people process such component information recently emerged from our lab. In four studies, Monroe and Malle (2014) assessed how people update initial blame judgments (made on the basis of verb-implied intentionality) in response to new information (explicitly mentioning intentionality, or good or bad reasons, or preventability). If people are guided by a desire to blame, they should persist in high initial levels of blame when they receive new mitigating information but should readily increase low initial levels of blame when they receive new aggravating information. Alternatively, people may update blame symmetrically in response to specific mitigating or aggravating information. In fact, this symmetry emerged in four studies, both when comparing all mitigating versus all aggravating cases and comparing, more specifically, new information about intentionality (present vs. absent), about reasons (good vs. bad), and about preventability (present vs. absent). Moreover, people's updated

blame judgments reached the same average levels as a control group that received all information at once and made a single blame judgment. Thus, we found no evidence for anchoring and insufficient adjustment of blame but strong evidence for differentiated updating as a function of key components of the Path Model: information about intentionality, reasons, and preventability.

### More Blame Motivation

A few other scholars have espoused models of motivated, biased moral judgment. Ames and Fiske (2013) recently proposed that people are so sensitive to intentional norm violations that they overestimate the harm that intentional acts produce, compared to unintentional events with identical consequences. In brief, people see intentional harms as worse even when, objectively, they are not. The authors explain this effect by postulating, like Alicke, a motivation to blame: "When people detect harm, they become motivated to blame someone for that harm . . . [and] seek to satisfy this motivation" (p. 1755). Critically, this motivation is said to bias people's judgments, in this case the assessment of the degree of harm that the norm violator actually caused. The authors show that intentional norm violations led to greater blame (compatible with the Path Model and many other models of blame) but also suggest that people's greater blame exaggerated their estimations of harm. The interpretation of exaggeration requires that harm was indeed "objectively" constant across intentional and unintentional conditions. We have reservations about this assumption, but instead of debating this issue we want to briefly discuss two questions about the motivation-to-blame construct in the studies.

First, "motivation to blame" was measured primarily like other researchers measure actual blame ("To what extent do you think Terrance deserves blame?"), so the evidence does not clearly speak to a *motivation* to blame but more to *judgments* of blame. And if judgments of blame need warrant, then participants may have offered perceived harm assessments as such warrant, with greater harm justifying greater blame. This does not necessarily imply that harm perceptions are biased, only that people infer them from base rates (in the real world, intentional events may generally produce more harm than unintentional events) and from the ambiguous stimulus material.

Second, if blame is an actual motive that can be satisfied, then learning that the harm-doer was caught, fired, and publicly blamed should decrease the motivation to blame. Goldberg, Lerner, and Tetlock (1999) called this "moral satiation." However, Ames and Fiske (2013, Study 3) found no satiation; people continued to see greater harm in the intentional than in the unintentional condition even when the

perpetrator was caught. This result further favors an interpretation of the data in terms of blame *judgments*, not *motivation*, because judgments should show no satiation—given whatever the agent did, he deserves a certain amount of blame, whether he has already received it or not.

Tetlock (2002; Tetlock et al., 2007) has argued that people adopt, under certain conditions, a “prosecutorial mind-set,” which fosters holding norm violators more culpable and punishing them more severely. Tetlock avoided the charge that “all blame is exaggerated” by identifying several variables that activate this mind-set: individual differences such as authoritarianism, emotions of moral outrage, attitudes favoring retribution, and beliefs about widespread and unchecked crime. If the evidence about a norm violation is ambiguous, Tetlock proposed, moral perceivers will take the opportunity to increase their punishment, relative to conditions under which the mind-set is not activated or the evidence is more clear-cut. Tetlock did not commit to any process model—for example, whether moral emotions come before causal and mental inferences, or whether judgments drive punishment or justify post hoc the desired level of punishment. All in all, the Path Model is compatible with this view, because the model allows for conditions under which processing is hampered or biased (see *Parsimony* section in Part 2), and its assumptions about cognitive processes are not contradicted by Tetlock’s model or findings. Tetlock also identified a number of mechanisms that help correct judgments potentially suffering from a prosecutorial bias, including information processing of the sort that the Path Model describes and responsiveness to social demands for warrant, which Tetlock and colleagues have called “accountability” (Lerner & Tetlock, 1999).

### ***Pervasive Morality***

Knobe’s (2010) analysis of the relationship between morality and social cognition is not directly a theory of blame but makes predictions that are opposed to the Path Model’s predictions. In particular, though Knobe conceded that judgments about causality and mental states guide blame judgments, he postulated an “initial moral judgment” (Phillips & Knobe, 2009) that precedes and directs this causal and mental analysis. Studies by Knobe and others suggest that, compared to positive or neutral actions, people judge negative actions as more intentional (Knobe, 2003), caused (Knobe & Fraser, 2008), and foreseen (Beebe & Buckwalter, 2010). The claim appears similar to Alicke’s, but Knobe considers these valence effects not to be *biases* but to demonstrate the pervasive role of moral considerations in the application of causal and mental concepts (Pettit & Knobe, 2009).

But questions arise about the evidence. For one thing, no study has measured the “initial moral judgments” that are claimed to affect intentionality and mental state inferences. And as long as studies are confined to text vignettes that present all information at once, such measurement is nearly impossible. In addition, few studies have assessed potential inferences people may draw from the critical manipulations. When studies did measure such inferences (e.g., about the agent’s desire or the action’s difficulty), valence effects on judgments declined or disappeared (Guglielmo & Malle, 2010a, 2010b). Last, many studies in this literature have capitalized on pragmatic demand effects typical for vignette studies (Adams & Steadman, 2004; Guglielmo & Malle, 2010a). For example, when a speaker asks a listener who “caused the problem” (Knobe & Fraser, 2008), the question is not aiming just at physics but at matters of fault; and when a speaker asks a listener whether an agent “knew about” his action’s negative side effect, the question is not aiming just at epistemology but at matters of obligation and counterfactual prevention.

It may appear that this is exactly Knobe’s point—that morality is intertwined with causal and mental concepts. But pragmatics is not semantics. If participants’ judgments vary by valence because they pragmatically read the experimenter’s communicative intention as inviting moral considerations, then this does not show that the semantics of epistemic and other mental concepts is fundamentally moral.

This distinction between pragmatics and semantics emerges when comparing experiments that vary the communicative demand put on participants. For example, in the well-known side-effect scenario (Knobe, 2003), a CEO knows that adopting a certain business program will harm the environment but nonetheless decides to adopt it because he “doesn’t care at all about harming the environment” and wants to increase profits. When participants are asked whether he harmed the environment intentionally, about 80% of participants check the box that indicates he harmed it intentionally. However, when participants don’t have to answer this forced-choice question but can select which of several descriptions is most accurate (i.e., The CEO willingly/knowingly/intentionally/purposefully harmed the environment), only 1% choose “intentionally” and 86% choose “knowingly” (Guglielmo & Malle, 2010a). People’s concepts did not change here; the communicative demands changed, and people’s judgments were sensitive to those demands.

We would like to mention, however, one consistent finding throughout Knobe’s experiments (and many other studies): People consider behavioral, causal, or mental information associated with norm violations more diagnostic than information

associated with nonviolations (cf. Reeder & Brewer, 1979; Skowronski & Carlston, 1989). Without entering a debate over the “true” diagnosticity of such information, we can confidently say that people’s cognitive system is keenly sensitive to norm violations (and not just to moral but also to nonmoral, even statistical violations; Guglielmo & Malle, 2010a; Pettit & Knobe, 2009; Uttich & Lombrozo, 2010). From our perspective, this underscores the enormous impact that the event detection phase has in the emergence of blame: It kicks the cognitive system into high gear, initiating the search for and processing of diagnostic information essential for arriving at blame. This information processing includes outcomes, motives, and character (Pizarro & Tannenbaum, 2012). Whether such processing, as a rule, is biased by motivational forces will continue to be debated.

### Social Intuitionism

Haidt’s (2001) social intuitionist model of moral judgment may seem, at first glance, to stand in direct contradiction to the Path Model of Blame. Haidt defined moral reasoning as “transforming given information about people in order to reach a moral judgment” (p. 818) but suggested that “moral reasoning is rarely the direct cause of moral judgment” (p. 815). The Path Model highlights the very elements and paths of such information “transformation” that generate blame judgments. However, Haidt’s theory is formulated for judgments of whether something is bad or wrong (type 2 moral judgments), not for judgments of blame (type 3 moral judgments). Indeed, studies that examined the intuitive/affective basis of moral judgments have always measured “wrongness”—essentially, people’s detection of norm violations (Haidt & Hersh, 2001; Wheatley & Haidt, 2005). The Path Model of Blame grants that people detect and evaluate norm violations quickly and often intuitively but holds that people *blame* an agent only after they process criterial information about causality, intentionality, and mental states. Such processing can at times be fast, especially when all the criterial information is available, and at other times more cumulative (Guglielmo & Malle, 2013). Either way, how people arrive at blame judgments is quite different from their “moral intuitions” about right and wrong.

### The Vexing Roles of Affective Phenomena

Many discussions over motivational forces in moral judgment appeal to affective phenomena—Alicke’s (2000) spontaneous evaluations are meant to be affective; Nadler (2012) suggested that character judgments influence blame through the perceiver’s

emotions; and Greene (2007) and Haidt (2001) regarded the fast, intuitive processes in moral judgments as primarily affective in nature. In fact, few scholars would doubt that affect and emotions play important roles in moral judgment. At the same time, empirical consistency and theoretical detail in research about these roles have been wanting (Huebner, Dwyer, & Hauser, 2009). The investigated phenomena range from raw affect to various specific emotions, especially anger and disgust, and the possible roles of these affective phenomena range from causing, to amplifying, to succeeding moral judgment (Avramova & Inbar, 2013; Horberg, Oveis, & Keltner, 2011; Pizarro, Inbar, & Helion, 2011). Some studies have examined emotions influencing type 2 (wrongness) judgments (David & Olatunji, 2011; Schnall, Haidt, Clore, & Jordan, 2008) or the other way around (Royzman, Leeman, & Sabini, 2008); others have examined type 3 (blame, responsibility) judgments influencing emotions (S. Graham, Weiner, & Zucker, 1997) or the other way around (Lerner, Goldberg, & Tetlock, 1998). Some studies have probed the impact of intentionality perceptions on emotion (Russell & Giner-Sorolla, 2011; Umphress, Simmons, Folger, Ren, & Bobocel, 2013); others looked at the reverse impact (Ask & Pina, 2011). Most important, however, the detailed psychological processes by which affective and cognitive phenomena might interact have not been systematically examined.

The Path Model, and especially its CIV process layer, can improve this situation. By demarcating different types of moral judgments, the model generates falsifiable hypotheses about the information categories (concepts) to which these specific moral judgments are sensitive; this then provides “locations” for potential interactions between emotions and the pertinent information processing (Chapman & Anderson, 2011). In addition, the model postulates three processes—the CIV triad—that operate at each information category: concept activation, information acquisition, and value setting. General affect or specific emotions can, in principle, interact with each of these processes. For example, being upset at the sight of an accident may lead to sharpened information acquisition for possible *agent causality*, admiring an agent’s prosocial character may preset the value of *reasons* to be *justified*, and a happy mood may lower one’s threshold of evidence for all components. At this point we can only speculate about how these processes interact, but we hope that the details of our model and a commitment to refined measurement approaches will provide answers in the future.

The Path Model of Blame also offers a reconciling position in the debate over early (often affective) and later (often deliberative) phases in moral judgment (Paxton, Ungar, & Greene, 2012). Rather than

contrasting affect and cognition and asking which one comes first, we rely on the distinction between early *event-focused* judgments and later *agent-focused* judgments (Malle et al., 2012; Monin, Pizarro, & Beer, 2007; Sher, 2006). People often experience negative affect toward norm-violating events along with a judgment of badness or wrongness. Event-triggered negative affect, however, is neither an emotion (which requires appraisals) nor a blame judgment (which requires causal and mental-state information). With further information processing, appraisals become available for emotions (Lazarus, 1984) and the perceiver's early affective response acquires meaning (Mandler, 1984). Thus, what distinguishes early evaluation from later blame is not a particular speed or mode of processing but the target of the processing—the event or the agent—and the particular information that is processed—violation of a norm or the agent's causality, intentionality, reasons, and capacity to prevent. Even this is probably too static a description, as information, evaluation, emotions, and judgments most likely build in iterative cycles and updates (Van Bavel et al., 2012).

#### Part 4: Applying the Model to Previous Results

We now describe how the Path Model of Blame accounts for a variety of findings in the literature—some puzzling, some problematic, some so basic that no theory can sidestep them.

##### Preventability, Not Controllability

In Weiner's (1993, 1995) theory, controllability and responsibility are prerequisites for moral judgments such as blame. These judgments vary depending on how controllable the *causes* of negative outcomes are. A student who fails a test is blamed if the failure was caused by his neglecting to study, which is a controllable cause. However, this leads to the counterintuitive prediction that *any* intentional action (which is, by definition, controllable) that causes *any* negative outcome leads to responsibility attributions, even when the action brought about the outcome in an unintentional manner. For example, at a party Jesse mentions the immaculate health of his 80-year-old father, which makes Gina very sad because her 80-year-old father just died. Jesse's utterance was certainly controllable, and it clearly caused Gina's sadness; but was Jesse therefore responsible for Gina's sadness and should one blame him? Most people would not. Rather than heeding the controllability of the *cause* of the outcome, people attend to the preventability of the outcome itself. Jesse neither knew about Gina's father nor was he capable of stopping Gina's emotion in its tracks, so

Jesse could not prevent Gina's sadness. This account is in the spirit of Weiner's theory, but it locates the critical criterion in the judged preventability of the outcome, not the controllability of its cause.

##### Repeated Behavior

Why are agents blamed more strongly if they repeatedly bring about the same or similar events (e.g., Robinson & Darley, 1995, Study 18)? Two cases need to be distinguished. In the first, the negative event is itself a series of behaviors (e.g., separately insulting three people at a party). Here, the evaluation is more negative because the norm violation is (summatively) more severe, and the perceived likelihood of intentionality is high because a pattern of repeated performance strongly suggests intentionality (Heider, 1958; Malle & Knobe, 1997b). The second case holds when an agent repeats a negative behavior after having been blamed the first time around. For repeated intentional actions, blame will increase because the agent is expected to have corrected any reasons that may have softened blame for the first-time offense (e.g., false beliefs, alternative goals). For repeated unintentional outcomes, blame will increase because, after the first offense, the agent is expected to have recognized her obligation and maximized her capacity to prevent the outcome.

The situation is different for cases in which moral perceivers evaluate an agent for a norm violation in one circumstance but know of the agent's "prior record" of having committed unrelated norm violations in other circumstances. This is essentially a case of character influencing blame, and we have discussed this complex relationship in Part 3.

##### Nonstandard Events

The most typical event that triggers blame judgments is a behavior that constitutes or brings about a norm violation. However, people blame agents for a variety of other events, including attempts, omissions, and cases in which a desired end is achieved by unexpected means. How does the Path Model handle such nonstandard events?

##### Attempts

People blame agents for their intentions, plans, and attempts; in fact, even for merely wanting or thinking about a harmful outcome (Guglielmo & Malle, 2012). Our model should apply to all such cases. To predict people's blame responses we must first ask exactly what was the detected norm-violating event. Suppose we observe a person holding a gun and entering a gas station, where he points the gun at the

cashier but is quickly overwhelmed by a nearby police officer. The event under consideration would normally be the *plan* or *attempt* to rob the gas station. Identified as such, the event's causal agency and intentionality information are already preset because agents are presumed to form plans intentionally. What is left for the perceiver to consider are the agent's *reasons* for attempting to rob the gas station (perhaps he was coerced into doing it; perhaps he hoped to pay the medical bills for his ailing wife). Thus, moral perceivers assign blame for an attempt in generally the same way as they assign blame for a completed action: by probing the agent's reasons for the action. But when we hold reasons constant, attempts and actions differ primarily in their initial severity of norm violation. The constitutive actions of *trying* to rob the bank usually violate fewer or weaker norms than the constitutive actions of actually robbing the bank (the latter involving far more manifest damage). Blame for attempts is therefore lower than blame for acts (e.g., Cushman, 2008; Robinson & Darley, 1995, Study 1).

### Omissions

Another nonstandard event that can receive consideration for blame is an omission to act. By definition, omissions are events that imply agent causality but leave minimal behavioral traces (DeScioli, Bruening, & Kurzban, 2011). Thus, event detection may be tentative or occur in steps: First, a negative outcome is found (e.g., a victim of a car accident dies), then an agent is identified who was copresent (another driver), which activates a prescriptive norm of helping that may have been violated. Search for intentionality information could then reveal that the copresent agent overlooked the injured person (unintentional event) or instead saw her and decided not to intervene (intentional event). If he truly could not see her, one might grant a lack of cognitive prevention capacity and therefore withhold blame. Some agents, however, have a strong obligation to *look for* potential victims when encountering an accident (e.g., police officers), in which case the person failed to meet this norm and deserves blame. If the agent actually decided not to intervene, the reasons for his decision will be critical in determining blame—for example, did he not want to get his suit bloody or did he help another crash victim? Thus, blame for omissions runs the course of the Path Model, but event specification may be slow or complex (unless it is formulated in language: "He did not extend his arm so the drowning victim couldn't grasp it").

In considering the well-known finding of omissions being blamed less than commissions (Cushman & Young, 2011; Spranca, Minsk, & Baron, 1991), we believe that there is no single factor that accounts for

the difference. The Path Model of Blame identifies three contributing factors. First, social perceivers may distinguish omissions and commissions by the norms these two actions violate. If there is a prescriptive norm to prevent a given outcome, then an agent's omission (not preventing it) will be readily detected as a norm-violating event—which we see in the blaming of agents who fail to report a presumed act of child molestation (Smith, 2011). Conversely, if there is no apparent norm to act preventively, an omission will not qualify as norm-violating.

Second, events of omission often have a more complex causal structure, which involves causal contributions from other agents or forces (Sloman et al., 2009). Researchers are careful in holding many things constant in their comparisons of omission and commission cases, but to hold the outcome constant across both cases, one must somehow implant an external cause into the omission story (otherwise the event would not happen). For example, in an oft-used case, a tennis player tries to poison his opponent during a joint dinner before the match by either (a) recommending a dish that contains a substance to which his opponent is allergic or (b) saying nothing when the opponent unwittingly orders the allergenic food himself. Even though the outcome is held constant (the opponent gets sick), perceivers' ascriptions of the agent's relative causal contributions will be different (smaller in the omission case, because the *victim* orders the food), which alters blame judgments (Cushman & Young, 2011).

Third, perceivers may be less confident about the agent's intentionality in the case of omissions because there is less evidence of an actual choice (DeScioli et al., 2011). Thus, the observed situation does not rule out that the agent failed to recognize the need to act, was indecisive, or had less committed intentions (Kordes-de Vaal, 1996).

### Vicarious Blame

A third nonstandard event stretches the notion of causality. Pet owners are sometimes blamed for damage caused by their pets; parents, for damage caused by their children; and company management, for accidents in the workplace. Such vicarious blame applies only when—following the unintentional path—obligation and capacity to prevent are plausible, which is typically guided by role and context. Parents have an obligation to prevent their child's transgressions, and employers have an obligation to prevent their workers' transgressions, but parents do not have an obligation to prevent their grown-up children's transgressions at work (Chiu & Hong, 1992). It might seem that vicarious blame violates the causality requirement in our model, because the one who is blamed (e.g., the pet

owner) did not directly cause the negative event (e.g., the dog biting a child in the park). However, people accept causation by neglect and thus consider the pet owner blameworthy for *allowing* it to happen that his pit bull roamed around the park and bit the child. Within counterfactual theories of causation, this is not a surprising claim: If only the owner had put the dog on a leash, it would not have bitten the child (Dowe, 2001).

### Wayward Causation

Sometimes agents perform actions, or achieve outcomes, in an unplanned, causally wayward manner. Imagine that George plans to stab his enemy to death. Now consider three ways in which he could accomplish this goal. In the first, George lunges forward and successfully kills his victim with the knife. In the second, before he lunges, George is hit by a jogger, falls forward, and thereby kills his victim. In the third, the victim sees the knife and is so scared that he has a heart attack and dies. Pizarro, Uhlmann, and Bloom (2003) showed that, in cases like the second and third—when the immoral act is committed in a causally wayward manner—people reduce blame. The authors suggest that current theories of blame “are unable to account for such blame reduction” (p. 653). The Path Model can. In all deviant cases, the actual immoral behavior is unintentional (in fact, the authors’ vignettes often marked this fact explicitly with words such as “accidentally” or “by chance”). At the same time, the offender had a full-blown intention to commit the act, and the desired outcome did occur. Thus, seeing the two cases side by side (in the studies’ within-subject designs), perceivers faced similar but distinct event structures: *intention + intentional action + outcome* versus *intention + unintentional behavior + outcome*. Perceivers are thus invited to assess the weight of the distinguishing middle element. Countless times in everyday life they have adjusted blame when an outcome arose unintentionally rather than intentionally; so, too, in these cases, they feel compelled to make an adjustment. The adjustment in Pizarro et al.’s (2003) studies was small because the highly immoral intention was present either way; but the adjustment is due to one critical difference: the perceived intentionality of the agent’s actual behavior.

Similar considerations explain Plaks et al.’s (2009, Study 1) pattern of results, which used the following wayward causal chain (originally devised by Chisholm, 1966): An agent plans to kill his uncle by hitting him with a car and either succeeds as planned or accidentally runs over a pedestrian, who turns out to be his uncle. Plaks and colleagues formulated the case in terms of “proximal” and “distal” intention. We interpret the study as

manipulating the intentionality of the critical behavior (causing a person’s death), so people judge intentionally killing the uncle as worse than accidentally killing the pedestrian while also incorporating blame for the original murderous intention in each case.

### Intervening Causes

A related challenge comes from cases in which a causal force *intervenes* between the agent’s behavior and the eventual outcome. For example, an agent tries to kill a victim and inflicts a gunshot wound; treated for the wound in the hospital, the victim dies of an allergy to a treatment drug. How much blame does the shooter deserve? Robinson and Darley (1995, Study 17) had participants assess criminal liability, but the results should generalize to blame. The most interesting variants of this case yielded the following results:

Case 1. A clear-cut intentional murder (the agent shot and *killed* the victim) received a liability rating of 9.9 (on a 0–11 scale).

Case 2. When the agent shot, *wounded* the victim, and the victim died of an allergy during the treatment of the gunshot wound, the rating was 8.8.

Case 3. When the agent shot, *missed*, and the victim decided to flee to avoid further risk, only to die in an *accident* 10 blocks from his house, liability was 7.4.

Case 4. A clear-cut failed *attempt* (the agent shot, missed, and the victim was unharmed) received a rating of 7.3.

To apply the Path Model, we need to precisely specify the judged events, and the experiment is set up such that some cases have two events—the agent’s action and the outcome caused by that action. In all cases, the agent *attempted to kill* someone, and when no real harm ensued (Case 4), the baseline level of blame was 7.3. Additional blame accrued in Cases 1 and 2, when the desired outcome obtained, but the action of *wounding* the victim (8.8) was blamed less than *killing* the victim (9.9) because it violated a less serious norm. In addition, Cases 2 and 3 involved events in the aftermath of the agent’s action that were unintentional. Thus, according to the Path Model, people considered whether the aftermath was caused by the agent and, if so, whether he was obligated and able to prevent it. Dying of an allergy to the gunshot wound (Case 2) is causally more proximal than dying in an accident (Case 3), and the agent did not have an obligation or capacity to prevent a new causal agent from hitting the victim. Thus, in Case 3 the agent is blamed only for the (failed) attempt to kill the victim,

with liability holding at 7.4, the baseline blame for the attempt alone.

We can take the same approach to a case by Cushman (2008, Study 3) in which an intervening cause appears (in italics):

Jenny wants to burn her lab partner's hand and believes that welding a metal will burn her hand. So she welds the metal, but her partner happens to let go and is not burned by Jenny. *Then the partner picks up a different piece of hot metal and is burned.*

Blame judgments were phrased as "How much blame does Jenny deserve?" which targets the *entire* event. Cushman found that, holding constant the agent's mental states (Jenny attempted to harm her partner), the agent received *less* blame when her partner picked up a different piece of hot metal and was burned (Variant 3) than when no injury occurred at all (Variant 1). This seemingly puzzling result emerges, we suggest, because people are asked to judge very different events: Variant 1 is Jenny's sole attempt (no harm caused), whereas Variant 3 is a multiagent composite of Jenny's attempt and her *partner's* causing her own injury. The partner's self-inflicted injury was in no way caused by Jenny, who therefore deserves no blame for it. Blame assigned to Jenny for the composite event (attempt plus injury) appears to be the average of the amount assigned to Jenny's attempt and zero (for partner's self-inflicted injury), resulting in a lower composite blame than the blame for Jenny's attempt by itself.<sup>7</sup>

Fincham and Shultz's (1981) study on blame in intervening cause scenarios provides another challenge the Path Model must meet. The authors constructed stories like the following: A primary agent wants to play a prank on a target person by hiding her ring in a shampoo bottle, but a secondary agent intervenes by using the shampoo bottle and flushing the ring down the drain, thereby causing more severe harm than the primary agent had ever intended. The authors showed that blame for the primary agent was lower when the *intervening* agent caused the harm intentionally or when the *primary* agent did not foresee the secondary agent's behavior.

Once more, the Path Model accounts for these results when we specify the precise events in question

and then probe the relevant blame components. Here the event was harm to the victim—set in motion by the primary agent's intention to play a prank on the victim but magnified in ways that the primary agent did not intend. Blame for the ultimate magnified harm therefore follows the unintentional path of our model, via obligation and capacity to prevent the harm. The control condition involved only the primary agent accidentally causing the magnified harm (the agent tried to hide the victim's ring in a shampoo bottle, but it slipped out of her hands and down the shower drain), and because the harm was preventable participants assigned a high mean blame of 7.9 (on a 1–9 scale). When the secondary agent *intentionally* caused the same harm, the primary agent was arguably neither obligated nor able to prevent the harm, whether she foresaw it or not (hence, mean blame dropped to 5.6). Nor was the primary agent obligated or able to prevent a secondary agent's *unforeseeable* behavior, whether intentional or not ( $M = 5.6$ ). Only when the primary agent could foresee that another person might unintentionally cause harm were any preventive steps obligatory and possible. When the primary agent failed to take such steps, she received a blame rating of 7.2, approaching the control condition's mean (though not quite, because another agent was causally contributing to the outcome).

## Summary

The Path Model of Blame clarifies a number of documented data patterns, including repeated behavior, attempts, omissions, and vicarious blame. If we properly specify both what the norm-violating event is and identify any preset values (e.g., agency for omissions, intentionality for attempts), then the model runs through the canonical conceptual structure and, depending on the particular values for the relevant concepts, predicts the proper blame judgments. The model also accounts for challenging wayward causation cases by highlighting the critical roles of event differentiation, intentionality, and of the specific combinations of prevention obligation and capacity. The model's predictions fit the data at an ordinal level, though our hope is that future model extensions will enable parametric predictions.

## Part 5: Blaming as a Social Act

One of the fundamental properties of blame is that it is both cognitive and social. So far we have focused on cognitive blame and the concepts and processes that support it; now we turn to social blame. The psychological literature is surprisingly limited on this topic, having made advances primarily on cognitive blame. We therefore rely here on relevant

<sup>7</sup>Cushman also asked a type 2 judgment: "How wrong was Jenny's behavior?" which focuses strictly on Jenny's action: welding the metal. Whatever happened afterward is not relevant to judging that event. Indeed, Cushman found that wrongness judgments were largely the same when no harm occurred (Variant 1) and when the partner inflicted an injury on herself (Variant 3). This result is consistent with our model because we conceptualize wrongness judgments as blame for the agent's action, which is held constant in the two scenarios, whereas blame is different between the scenarios because it covers both action and outcome.



philosophical and sociological literatures and extensions of our cognitive model of blame to the social level.

Regulating behavior is a core property of social blame. But by criticizing norm violations, acts of blame devalue the blamed agent. To minimize the potential cost of such devaluing social blame is itself regulated by social norms (Bergmann, 1998; Coates & Tognazzini, 2012b). If social perceivers harbor a desire to blame (Alicke, 2000; Ames & Fiske, 2013; Tetlock et al., 2007), then norms of social blaming would limit when this desire can be publicly satisfied. Some of these norms are culturally and historically variable, including expectations about who is allowed to blame whom, in what contexts, and for what offenses. There are even highly local norms about how often and in what tone social blame is expressed—which everybody knows who had opportunity to compare, say, an upper-class British family and an Italian family (cf. Corsaro & Rizzo, 1990). But elucidating social blame requires us to focus on the structure of social blame that transcends specific local norms. To do so we first situate the phenomenon of social blame within related public acts of moral criticism and then turn to its fundamentally communicative nature.

## Blame and Other Acts of Moral Criticism

### Social Acceptability

One attempt to organize the many forms of moral criticism is to ask how socially acceptable they are. Voiklis, Cusimano, and Malle (2014) elicited acceptability judgments from a group of participants who read 28 abstract action descriptions (“He [verbed] her for the bad thing she had done”), where each of the action description used a different verb of moral criticism. A second group of participants indicated how similar each verb phrase was to the standard phrase “He blamed her for the bad thing she had done.” The

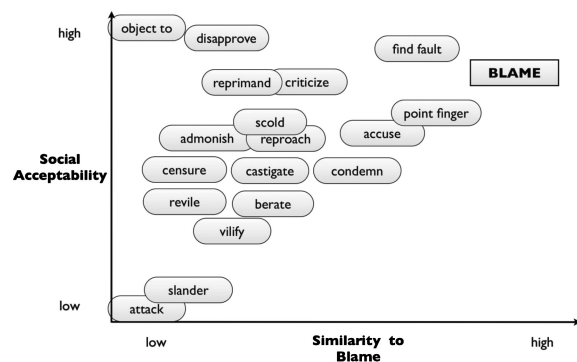


Figure 4. Social acts of moral criticism ordered along the dimensions of social acceptability and similarity to blame. Note. Based on judgments averaged across separate groups of participants.

results in Figure 4 represent a streamlined depiction of Voiklis et al.’s data (showing 17 of the 28 verbs). Blame emerges as one of the most accepted forms of moral criticism, along with *finding fault* and *pointing the finger*. The acts that are least socially acceptable and most unlike blame are *attacking*, *slandering*, and *vilifying*. These results mirror those of Alberts (1989), who found in interviews with couples that by far the least desired forms of complaint behavior were yelling and personal attacks whereas the most desired ones included rational, calm, constructive criticism.

### Emotion and Thinking

Taking up this contrast between yelling and calm criticism, another way of grouping acts of moral criticism is within a two-dimensional space of emotional intensity and thoughtfulness. The plotted verbs of the blame family in Figure 5 show again data from Voiklis et al. (2014). Participants judged either how intense the emotion was that the perceiver must have felt or whether the action sounded more impulsive versus more thoughtful. Acts of blaming were judged to have at least moderate thoughtfulness and lower emotional intensity, in the neighborhood of *rebuking*, *reproaching*, *accusing*, and *scolding*.

We therefore conclude that social blame is an acceptable act of social regulation, affective enough to signal seriousness (McGeer, 2012a) but favoring thought over emotional intensity. This pattern allows blame to be a deeply communicative act, which we explore next.

## The Communicative Structure of Blame: Persuasive Blaming

Social blame is by nature communicative—both when the blamer directly addresses the norm violator (second-person blaming) and when the blamer talks to others about the norm violator (third-person blaming). We begin with the communicative processes

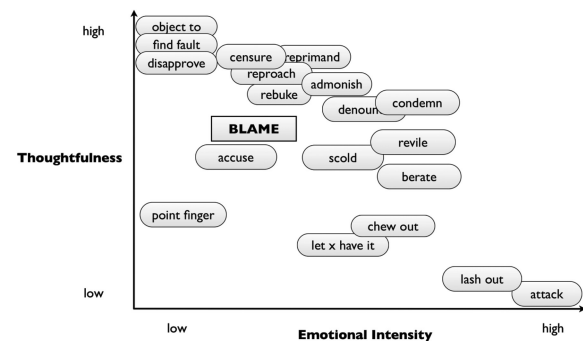


Figure 5. Acts of moral criticism within the space of emotional intensity and thoughtfulness (vs. impulsiveness). Note. Based on average judgments of two groups of participants.

that are unique to second-person blaming—in what we call *persuasive blaming*.

Persuasive blaming is perhaps the oldest form of human moral regulation. In the 40 to 80,000 years before human settlements (about 10,000 BCE), humans lived in small bands of 25 to 50 in nomadic life styles (Boehm, 1999; Knauff, 1991). We know this partially from archeological finds (Bandy, 2004; Enloe, 2003; Taçon & Chippindale, 1994) but predominantly from ethnographic research of hunter-gatherer societies over the past 100 years (e.g., Leacock & Lee, 1982; Lee, 1972; Lee & Daly, 1999; Service, 1966; Wiessner, 2005; Woodburn, 1982). From this we can infer that most hunter-gatherer communities were highly egalitarian, with the exception of some gender and age differences in social influence and decision making (Carling, 2000). There was no one centralized ruler, lawmaker, or judge; leadership was provided by different members for different tasks (Service, 1966). Everyone knew each other, and maintaining relationships was critical to survival of the individual and the group.

In such communities, sanctioning and conflict resolution were interpersonal. Most norm violations occurred publicly because community life was inherently transparent (Silberbauer, 1982; Wilson, 1988, Chapter 2). Community members responded to such violations with criticism, ridicule, or temporary ostracism rather than with physical punishment or permanent banishment (Boehm, 1999). In conflicts, the wronged party would point out the offender's norm violation, and the two parties negotiated mild punishment or compensation to restore social equilibrium (Rouland & Planel, 1994, p. 167). When no satisfaction was reached, cases moved before the group where an arbiter or elder would make a recommendation for sanctions or restitution (Pospisil, 1971); but it was up to the involved parties to follow the advice and find reconciliation.

These practices of moral regulation through negotiation and persuasion also characterize many of today's instances of social blame. Blame demands a response (Drew, 1998; McGeer, 2012a; Newell & Stutman, 1991; Shoemaker, 2012), and in particular an interaction between the blamer and offender to repair their strained relationship (Bennett, 2002; Goffman, 1967; Walker, 2006). Even the legal system—after centuries of institutionalized, often brutal methods of punishment—has rediscovered communicative forms of regulation in the form of restorative justice procedures (Kuo, Longmire, & Cuvelier, 2010; Rossner, 2011). In these procedures, offender and victim—even though they are typically strangers—rebuild the symbolic relationship that everybody has, or should have, with their community.

Although empirical data are in short supply, work in philosophy, sociology, and communication

suggests several preconditions for persuasive blame to be successful.

- *Joint attention.* The blamer grabs the offender's attention, perhaps through a clear display of emotion (McGeer, 2012a), or perhaps through a direct statement of the violated norm (Drew, 1998).
- *Communication.* Blamer and offender communicate about the norm violation (McKenna, 2011; Pearce, 2003), and the offender receives an opportunity to provide, if appropriate, relevant causal-mental information. This information might change the blamer's social-cognitive information base, and thus his warrant, for the specific degree of assigned blame.
- *Delivery.* As mentioned earlier, Alberts (1989) found that yelling and personal attacks were the least desired expressions of complaints in couples, whereas partners welcomed rational, clear, and constructive criticism. It would seem obvious then that persuasive blaming holds the greatest promise when blame is delivered with low emotional intensity and high thoughtfulness—producing the most socially acceptable moral address (Voiklis et al., 2014).
- *Shared values and community.* The blamer does not simply condemn the other person's behavior but focuses on the shared values or personal expectations that have been violated (Walker, 2006), with the hope that the offender recognizes the wrongness of her actions (Duff, 1986b; Schmitt, 1964). To engender this insight the blamer must treat the offender as a member of the community (Bennett, 2012) who deserves respect and the presumptions of autonomy and rationality (Duff, 1986a; Holroyd, 2007; Wolf, 2011). Under these conditions, the offender may recommit to the very values she had violated (Metts, 1994).
- *Repair.* The damage to the parties' relationship must be repaired through the violator's adequate response to the blamer's demand (Bennett, 2002; McGeer, 2012b; Walker, 2006), such as admission, acceptable justification, sincere remorse and apology, and sometimes restitution. When such a response is not forthcoming, regulation of social relationships fails (Laforest, 2002). Even revenge and punishment do not succeed without the offender offering at least some acknowledgment of the violation (Carlsmith, Wilson, & Gilbert, 2008; Gollwitzer, Meder, & Schmitt, 2011). In extreme cases, a

justification or apology occurs preemptively—even before a complaint is voiced (Schegloff, 2005).

- *Social cognition.* Social-cognitive processes contribute to blame's regulatory function by targeting, through persuasive communication, the psychological basis of an agent's future behavior: the *reasons* for acting one way or another. In episodes of persuasive blaming people present reasons to the offender for why she should have acted differently at the given occasion and thus reasons for why she should take an alternative action at similar occasions in the future. Communicating blame thus directly influences the offender's decision process about not committing the norm violation in the future (G. P. Miller, 2003). Moreover, by providing reasons to the agent in an attempt to influence this decision process (rather than, for example, physically impeding the agent's behavior), the blamer communicates a conviction that the agent is competent to follow norms on her own accord and to change her behavior (Holroyd, 2007).

### Third-Person Blaming

The constructive features of persuasive blaming are necessarily absent in third-person blaming—which is blame addressed to other observers in the offender's absence.<sup>8</sup> With little chance of (or interest in) reforming the offender, such blaming serves to express the blamer's emotions, reassert the violated norms, and seek validation for those norms (Drew, 1998; Duff, 1986a; Pearce, 2003). Audiences of third-person blaming often affiliate with the blamer and thus affirm shared norms and provide legitimacy for the complaint (Laforest, 2009). Because the audience often joins forces, third-person blaming sometimes represents a first step toward socially excluding the offender (Kurzban & Leary, 2001). But all of this is possible only if the blaming can be supported by appropriate warrant. Indeed, sociolinguistic research shows that third-person blaming episodes are more elaborate than second-person blaming episodes (Dersley & Wootton, 2000; Drew, 1998; Traverso, 2009). The blamer typically describes in detail the context of the transgression, the specific transgressive act, and sometimes ends the grievance with a graded affective report ("I was so angry"; "that teed me off"; Drew, 1998, pp. 309–311). The desire to build an

alliance and the pressure to provide warrant may also make people vulnerable to exaggerating the informational elements that normally warrant blame, such as motive and degree of harm (Ames & Fiske, 2013; Haidt, 2001).

### The Darker Side of Moral Criticism

In practice, things don't always go so well in moral communication. The blamer might choose an act closer to the lower right corner of Figure 5, high in emotional intensity but low in thoughtfulness. And rather than responding to the content of the blaming, the offender may mirror the emotional intensity of the blamer's expression, with escalation following suit (as, e.g., confrontations in traffic amply illustrate). Furthermore, targets of blame easily get "defensive" and rather than showing insight, remorse, and making amends, they often reject the criticism (Dersley & Wootton, 2000; Laforest, 2002). Occasionally they even attack the blamer and find something for which to criticize her in return, be it the blaming act itself, a lack of warrant, her standing, or some other behavior worth criticizing. Such patterns of complaint-countercomplaint are particularly common in dissatisfied couples, relative to satisfied couples (E. J. Thomas, 1977). Blamers don't respond too well, of course, to countercomplaints, because they thwart her goal to "right" the offender and any hope for repair (Alberts, 1989). If the blamer then contests the offender's rejection of the blame, conflict is likely (Dersley & Wootton, 2000; Laforest, 2002). In such cases the constructive function of blame as relationship repair has not been achieved.

The constructive function of blame is also likely to fail when the value of repairing the relationship is missing: between strangers, who don't have such a relationship. Outside of court-appointed arbitration and restorative justice procedures, there is little pressure to communicate, persuade, repair, and find common ground with a stranger. Instead, moral criticism becomes akin to road rage, an episode of Jerry Springer, or hateful anonymous comments on the internet (Santana, 2012). It isn't that there are no longer any norms in stranger interactions; it's that people are far less motivated to acquire sufficient information and are far less likely to be called on for the lack of warrant in their judgments. When such lack of warrant becomes obvious, most people are perfectly capable of switching back into the civil mode. Just observe the screaming driver who suddenly notices that the other driver whom he had reviled is actually in distress or, worse yet, turns out to be his neighbor. Self-regulation immediately takes the upper hand, showing the powerful impact of cognitive appraisals on emotions and the impact of norms on acts of blaming.

<sup>8</sup>This is distinct from "third-party" blame or punishment (e.g., DeScioli & Kurzban, 2009; Fehr & Fischbacher, 2004), which refers to sanctions by an uninvolved party (i.e., not a victim), irrespective of whether the perpetrator (or even an audience) is present.

A recently formed norm of blaming is entailed by the expression “(playing the) blame game,” which emerged in 1958, according to the *Oxford English Dictionary* (Simpson & Weiner, 1989). At its core it describes the activity of assigning blame, finding fault after a negative event has been discovered; but it clearly is an undesirable variant of blame: “the game itself is blameworthy” (Robbins, 2007, p. 140). It often involves multiple people blaming each other—“pointing fingers” at multiple candidate targets. The undesirable nature of the game is that its players consistently accuse others of wrongdoing while deflecting or denying their own wrongdoing (Furlong & Young, 1996; Knobloch-Westerwick & Taylor, 2008). Detached observers, who criticize the players of the blame game, want one or more of those involved to “take responsibility” or “shoulder the blame.” Neither the detached observers, however, nor the players of the blame game operate without reflection, willy-nilly picking targets of blame. They all *argue* for their accusations and defenses, trying to offer warrant for their blame by selecting the familiar concepts and contents that the Path Model of Blame identifies—causality, intentionality, reasons, and so on—this time, however, with sloppy information processing, or in the form of outright lies.

Frequent unjustified blaming may signify a defective relationship (Fincham, Beach, & Nelson, 1987). Matters become worse when a blamer not only criticizes the other for having done something norm-violating but generally rejects and invalidates the offender. Here, the moral critic has dispensed of all argument and reform and expresses hateful derogation—“one must see and spoil the other, one must disfigure them” (Furlong & Young, 1996, p. 194). Such acts of hate, however, should be distinguished from blame. People consider such acts to be unjust precisely because they wholly ignore—and refuse to probe—the foundational questions of blame: Was the agent causally involved? Did he act intentionally? Could he have prevented the outcome? The evolution of legal systems may in part be a collective attempt to avert the most hateful and unfair moral sanctions—an attempt to establish binding norms of blaming.

When one group is in power, however, it can rewrite the norms of moral criticism and single out certain others as targets of blame (Douglas, 1995). Selecting such “scapegoats” can in fact increase the coherence of a group and aid in the collective endeavor of accounting for negative events (Treichler, 1999). One of the most cruel examples is the Nazi propaganda to blame Jews for the economic crisis and cultural “ills” of Germany in the 1930s. This propaganda led both to increased group coherence (nationalism and wide support for the Nazi party) and to the brutal escalation of legalized social exclusion all the way to genocide. Of importance, the

propaganda claimed specific causal, even intentional, contributions of Jews to the society’s woes. It was not just an irrational lashing out stemming from negative affect; on the part of the propagandists, it was a systematic “argument” in line with the informational and conceptual components of blame, and it had lasting effects on the population’s emotions, judgments, and actions.

## Blame Management

Because blame imposes social and psychological costs on the person blamed, quite some effort goes into managing and curtailing moral criticism, as noted in a voluminous literature (e.g., Benoit, 1995; Cupach & Metts, 1994; Goffman, 1967; Scott & Lyman, 1968; Semin & Manstead, 1983; Snyder & Higgins, 1988; Weiner, Figueroa-Munioz, & Kakihara, 1991). Dersley and Wootton (2000) reported that 95% of second-person complaints (many of which can be classified as blaming) are to some degree contested, and Alberts (1989) found that denials and justifications make up 65% of spousal responses to their partner’s complaints (a reasonable proxy for blaming). The Path Model of Blame specifies what information is contested in such blame-managing responses—namely, the very same information that normally grounds a blamer’s private judgment of blame in the first place and that is meant to warrant the corresponding act of blaming. If this information base can be corrected or undermined, then blame is less warranted and may be reduced or even revoked.

Research on blame mitigation has catalogued various physical, psychological and social factors that may reduce blame (Alicke, 1990; Heath, Stone, Darley, & Grannemann, 2003), but it has lacked a strong theoretical framework. Some models of moral judgment have explicitly integrated mitigation (e.g., Alicke, 2000; Weiner, 1995) but often in the general sense of negating blame-relevant information that normally guides moral judgment. Exactly what types of information can be negated is less clear. For example, a claim of “uncontrollable” or “external” causes may mitigate blame for unintentional negative events, but it won’t work for intentional actions, which are by definition controllable and internal to the agent. Moreover, several classifications of blame-mitigating attempts have been so fine-grained, with more than 20 different types (e.g., Scott & Lyman, 1968; Tedschi & Reiss, 1981), that no integration into a comprehensive model has occurred.

The Path Model of Blame provides an organizing framework for this literature because mitigation strategies can be directly derived from the conceptual structure of blame (Figure 6). Every information node that normally builds a blame judgment can be denied, questioned, or revised. For example, if

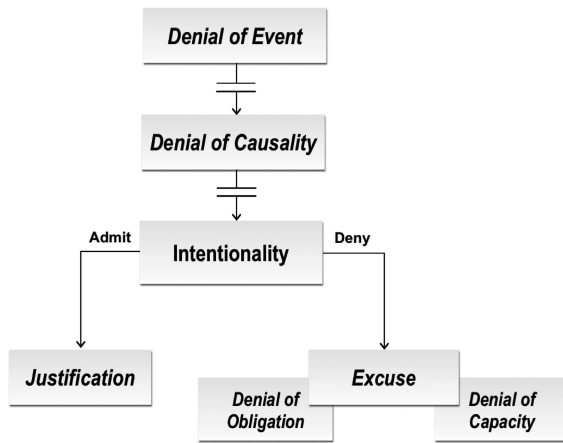


Figure 6. Blame mitigation strategies derived from the Path Model of Blame.

somebody causes a traffic accident by hitting the car next to him he might explain his behavior by saying “You were right in my blind spot” (unpreventable), “I didn’t mean to” (unintentional), or “I was trying not to hit the little girl in the crosswalk” (justifying reason). And just as intentionality carves two separate paths of information search en route to blame so it opens two major paths of information revision en route to blame mitigation—providing *excuses* for unintentional events (primarily, negating obligation or capacity) or *justifications* for intentional actions (primarily, reason explanations).

We now examine these mitigation strategies in more detail.

### Denial of Event

The defender’s most radical option is to deny the norm-violating event—either by denying the event’s existence (“It didn’t happen”) or by denying the legitimacy or applicability of the norm that was allegedly violated (Metts, 1994; Newell & Stutman, 1988). If either of these claims is evidently true, it would keep the defender blameless, but strategic event denials without good evidence rarely succeed (Dersley & Wootton, 2000). The offender can also try to dispute the *nature* of the alleged norm-violating event (e.g., “I’m guilty of sex and contributing to the delinquency of a minor, but not rape”; Scully & Marolla, 1984, p. 537) or claim that the event itself is not norm-violating (“Around here almost everyone has taken some kind of a bribe at one time or another”; Riordan et al., 1983).

### Denial of Causal Agency

If the event itself is acknowledged, the defender can most quickly protect against blame by denying

causal agency. Such denial may focus on the *agency* element by providing evidence that, even though the person was causally connected to the event in question, he did not meet moral eligibility standards (e.g., due to age or mental status; Alicke, 1990; Fincham & Roberts, 1985). Alternatively, denial may focus on the *causality* element by providing evidence that, even though the person met moral eligibility standards, her causal connection was negligible or absent (e.g., “I didn’t dent the car”; “I was somewhere else that night”). The no-agency defense, if credible, can completely avert blame but carries the cost of designating the agent morally ineligible and thus at lower standing in the social community. The no-causality defense can be tenuous because causal connections come in many degrees and forms, and an agent’s mere presence at the scene may preserve suspicions of his involvement. In particular, because of the concept of *allowing* causation, an agent may be blameworthy for failing to meet her obligation to prevent a negative event even if she did not directly cause it.

If the agent’s causal involvement is evident, the next options are to deny intentionality and offer *excuses* for the purported unintentional event (“I couldn’t have known”; Markman & Tetlock, 2000) or to admit intentionality and provide *justifications* for the intentional event (Gollan & Witte, 2008). The Path Model characterizes justifications as socially acceptable reasons for intentional actions and excuses as unpreventable causes for unintentional events. This characterization (paralleling Fillmore’s, 1971, which was derived from linguistic data) provides a strong theoretical foundation for what justifications and excuses are and resolves previous disagreements over the best way of distinguishing the two (e.g., Greenawalt, 1984; Husak, 2005; Semin & Manstead, 1983).

### Justifications

Justifications as reasons come primarily as beliefs or desires (Malle, 1999, 2011). In their justifying use, beliefs can be mistaken but have to be sensible (e.g., that one’s life is in danger), while desires have to be socially desirable (e.g., to save a patient the doctor amputates a limb). In both cases, justification is a continuous value, varying with the degree of credibility and cultural acceptability of the provided reasons (e.g., Cohen & Nisbett, 1994) and with the extremity of the norm violation (Robinson & Darley, 1995). Particularly harmful actions (e.g., killing) require stronger justifications (e.g., self-defense)—that is, desires with great social value or beliefs that are well founded in reality. The desire reason “I just wanted to scare her a little” may suffice to justify telling a lie but not to justify committing a rape (Scully &

Marolla, 1984). There is some evidence that belief reasons outperform desire reasons in eliciting an audience's blame mitigation (Malle & Nelson, 2006), and in studies of people's attempts to self-exonerate acts of violence, belief reasons seem to dominate: "people have to be put in their place"; "it was my job to punish"; "it won't hurt them too bad" (Bandura, Underwood, & Fromson, 1975).

Justifications also apply to nonstandard cases such as actions under extreme social pressure or duress (e.g., committing a crime under threat to one's life). The action (committing the crime) is intentional; however, because the agent had severely constrained options, and none of the alternative options was acceptable, the community acknowledges that the agent behaved like any reasonable person would and therefore reduces blame (Reeder, Monroe, & Pryor, 2008; Woolfolk, Doris, & Darley, 2006). Psychologically, people may simulate the actor's distressing decision conflict and, sensing that the only option for them would be just the one the agent chose, they find that the agent acted with justified reasons.

### Excuses

When intentionality is ambiguous agents may be able to deny that an event was intentionally caused. Indeed, much of the literature on excuses has focused on denying intentionality (De Brigard, Mandelbaum, & Ripley, 2008; Semin & Manstead, 1983; Tedeschi & Reiss, 1981). Although the results of these studies are not entirely consistent, several of them find that the most effective blame-mitigating factors are those that alter or bypass the normal intention formation or choice process (e.g., diminished capacity, psychological disturbances, brain abnormalities).

Yet denying intentionality by itself rarely achieves blame mitigation. Intentionality bifurcates perceivers' further processing of norm-violating events; it does not terminate the process of blame. Denials of intentionality shift a perceiver's focus from mitigating by justification (along the intentional path) to mitigating by excuses (along the unintentional path). Blame for an unintentional event may still be high if the agent should and could have prevented it but did not take preventive steps; so the defender must convince the audience that he either didn't have an obligation or didn't have the capacity to prevent the event or, in fact, took preventive steps.

The tactic of denying an obligation to prevent the negative event will rarely be successful. Many moral proscriptions explicitly obligate community members to prevent a certain type of event from occurring (whether that occurrence is intentional or

unintentional). If an agent denies such an obligation she would thereby either exempt herself from the community's system of moral norms ("Why should I have to worry about that?") or question that system altogether ("What's so bad about *that*?"). Excusing by denying an obligation to prevent may be most successful if an agent's specific role legitimately exempts her from the obligation in question (e.g., "I'm just a programmer; I'm not responsible for monitoring the company's food safety practices").

The tactic of denying a capacity to prevent the negative event may appeal to cognitive limitations (e.g., "I could not see it") or physical constraints (e.g., "I couldn't do anything about it"). Among cognitive limitations, excusing by simple ignorance ("I had no idea this would happen") is popular (Markman & Tetlock, 2000), but often insufficient. To reduce blame—say, for an unintended side effect—an agent must also demonstrate that she made some effort to acquire information about possible side effects (Alicke, Buckingham, Zell, & Davis, 2008); otherwise the excuse can easily be rejected by saying, "You should have known that." Physical constraints are also most effective if they show themselves in an agent's trying but failing to prevent the event in question or in a patently insurmountable obstacle ("I could not stop because there was ice all over the road").

### Reconciliation

Blame management through mitigation, sometimes truthful, sometimes not, is a fundamental property of social blame. For this process, the cognitive structure of blame provides an organizing framework. There are, of course, steps after blame, and thus beyond the Path Model, that do not primarily involve mitigation but rather reconciliation, such as admission, remorse, apology, and restitution. These steps have the power to successfully repair relationships, often through the moral perceiver's forgiveness (Allan, Allan, Kaminer, & Stein, 2006; McCullough, Kurzban, & Tabak, 2013).

### Limitations

We have introduced a new theory of blame. We define blame as a unique moral judgment that has four properties: Blame is both cognitive and social, regulates social behavior, fundamentally relies on social cognition, and requires warrant. At the heart of the theory lies the Path Model of Blame, which identifies the conceptual structure in which blame judgments are embedded and the psychological processes that generate such judgments. In addition to discussing blame as a cognitive process we have also explored

blame as a social act, a phenomenon that has received far less attention in the psychological literature. Ongoing and future research will have opportunities to address some of the present limitations of this theory.

First, we cannot be sure that the Path Model's posited conceptual framework is complete—that there is no other information condition that influences blame. Theories grow with research they spark, so we expect that any significant omissions will soon be discovered. Evidence is also still needed on specific exclusionary claims of the model, such that wrongness judgments are equivalent to blame judgments for actions or that responsibility judgments make no independent contribution to blame.

Second, we have adopted a pluralism about modes of processing en route to blame judgments, arguing that those processes can be automatic or controlled, unconscious or conscious (Kruglanski & Orehek, 2007; Mallon & Nichols, 2011). Our theoretical commitment is that the cognitive path to blame is instantiated by an integrated set of information conditions, not by any particular processing requirements. Nonetheless, future research may be able to clarify whether some concepts (and their value settings) favor one processing mode over another.

Third, we have not yet sharply delineated the role and impact of affect in the information processing chain. Affect will often enter the event detection phase as negative evaluation. Whether affect is powerful enough to make people skip or markedly distort information processing steps is an open empirical question. To make a strong case for the power of affect, researchers must independently vary both affective and information parameters. The mere impact of an affect manipulation on levels of blame (which extant studies have demonstrated) does not address the actual process that underlies such an impact. Our model specifies the information processing steps that need to be manipulated or measured for the data to speak cleanly to this issue.

Fourth, some may consider the Path Model too “rational” a model of blame. However, the constraints that the perceiver obeys are information integration constraints, not rationality constraints. People undoubtedly can ignore information, make false assumptions, or blame to satisfy a strategic goal. Our claim is that people's blame judgments conform to the specified concepts of the Path Model, not that people always process information about these concepts in an objective or unbiased manner. Socially expressed blame, in particular, can deviate from the information structure of private blame—though it cannot deviate too much or too often because people do warrant, defend, and contest such blame judgments with precisely the kind of information that

normally guides private judgments. The Path Model of Blame accounts for most blame judgments most of the time, and deviations from the model are expected just like for any other psychological theory. However, improvements can be made to the model by identifying the conditions and extent of such deviations.

Fifth, our analysis of blame as a social process, though guided by the Path Model, went far beyond current evidence. We hope that readers will agree that social blame is worthy of increased empirical research, which will in turn refine the social layer of our theory of blame.

Sixth, a major limitation of this and all extant models of moral judgment is that they do not generate any quantitative predictions. We hope to expand the Path Model in ways that will allow such predictions. The simplest approach would be a multiplicative model of all the conceptual nodes as variables: initial event evaluation; agent causality (0 or 1); causal contribution (up to 100%); and, for intentional behaviors, reasons (scaled for degree of justification). But such a model fails to represent the dynamic order of processing that, we have argued, often guides blame judgments—for example, if agent causality = 0, no other variables need to be computed. Moreover, a detailed model would also integrate the “microprocessing” that forms the CIV layer. A related intriguing question is how people actually scale blame judgments in real life. In an experiment (and a test of a quantitative model), participants can be asked to use rating scales; but in everyday moral judgments, the situation is quite different. People scale the intensity of their blame by words, affective expressions, and choice of social actions, none of which are easily parameterized. Nonetheless, the eventual goal of a theory of blame must be to solve these problems and offer fine-grained quantitative predictions.

### Funding

This work was supported in part by the National Science Foundation (Grant BCS-0746381), the John Templeton Foundation/FSU Research Foundation (Subaward SCI05), and the Office of Naval Research (Award N00014-13-1-0269).

### Note

Address correspondence to Bertram F. Malle, Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, 190 Thayer Street, Providence, RI 02912. E-mail: [bertram\\_malle@brown.edu](mailto:bertram_malle@brown.edu)