Psychological Inquiry, 25: 251–260, 2014

ISSN: 1047-840X print / 1532-7965 online
DOI: 10.1080/1047840X.2014.913379

Routledge
Taylor & Francis Group

REPLY

Paths to Blame and Paths to Convergence

Bertram F. Malle
Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, Rhode Island

Andrew E. Monroe
Department of Psychology, Florida State University, Tallahassee, Florida

Steve Guglielmo
Department of Psychology, Macalester College, Saint Paul, Minnesota

The past 10 years have seen an unprecedented rise in research on moral psychology. The thoughtful and creative comments on our target article illustrate the vibrancy of the field and the many open questions that are ripe for investigation. Reading these commentaries, we are heartened by the balanced engagement with both theory and evidence and the convergence of perspectives.

In discussing some of the major themes in the commentaries we first return to the question of what blame really is and how we can most fruitfully carve out the boundaries between blame and related phenomena. Then we take up the phenomena of victim blame and self-blame as important applications of our model and highlight the important explanatory work that the notion of *presets* plays. Next, we examine the early and highly consequential process of *event detection*. Finally we address the greatest challenge to our model: the role of motivational biases in the emergence of blame and how the model accounts for such biases. Throughout, we promote methodological standards we believe must be heeded for significant progress in understanding the psychology of moral judgment.

What Blame Is

One of the burdens of psychology is to investigate phenomena that are labeled by everyday terms. These terms are not designed for science but for maximal efficiency and flexibility of communication, permitting multiple meanings, vagueness, and other irritants for any scientist. Nonetheless, those everyday terms often pick out phenomena of fundamental importance to human experience and action; and as psychologists we surely want to study those phenomena. To build scientific theories we therefore need to find a compromise between using the terms in ways that still denote the phenomena people care about while also sharpening those terms and delineating the objects of scientific interest. Some of the commentaries explored the boundaries of blame for different targets, in different contexts. From the perspective of our theory, here is how we suggest to draw those boundaries.

Blaming Whom?

Our investigation of blame was focused on a phenomenon of moral criticism that we defined by several properties (pp. XX): (a) Blame is a cognitive judgment and a social act, (b) blame's function is social regulation, (c) blame relies on social cognition, and (d) blame requires warrant. These properties are meant to separate off everyday uses of the word "blame" that are partial or metaphorical. Do people literally blame fate, gods, sharks, and hurricanes (Alicke, DeScioli, Goodwin)? In all these cases we see partial overlap with blame as we investigated it, but there are also notable differences.

Blaming deaths on hurricanes or other natural phenomena is primarily a causal judgment; social cognition and social regulation are arguably absent. If one considers this judgment "blame," then every causal judgment of negative events would be "blame."

Blaming fate and gods may, depending on one's conception of fate or gods, involve not only causality but also social cognition (agency and intentionality ascriptions), in which case it comes closer to prototypical blame. Whether such blaming functions as "regulation" depends further on one's assumptions of the persuadability of fate and gods. If they are persuadable (and at least in some religions, God is), then we expect that this supernatural blaming will take

into account the agent's reasons. The phenomenon of blaming animals for damage they caused will also vary as a function of the perceiver's conception of the animal's capacities. Blaming the shark for three deaths is likely to be taken as a purely causal statement, except perhaps by the animal trainer who had a relationship with the shark. Likewise, blaming one's pet has the potential of coming close to prototypical blame if the owner actually applies social-cognitive inferences (of intentionality) and performs social-communicative acts of blaming (not just punishing) for the purpose of behavior regulation.

Pizarro raises the interesting question of whether humans would be willing to blame robots and other artificial machines. We very much expect it. Humans are perfectly happy to interact and communicate with a robot, even have a relationship with it (Kahn et al., 2010; Scheutz, 2012), and to the extent that people use their social-cognitive framework to interpret the robot's behavior—hence ascribe intentionality, mental states, and the capacity to respond to persuasion—they are likely to blame the robot as well (Kahn et al., 2012; Monroe, Dillon, & Malle, in press).

In sum, the key prediction from our theory is that, *if* all four defining properties of blame are met, we expect that the blamer will process information along the conceptual-cognitive paths we have outlined (such as taking into account God's reasons).

Blame's Relatives?

Sheikh and McNamara (this issue) suggest that "certain emotion categories are in fact types of blame" (p. 241) such as contempt, disgust, outrage, and anger. We are not sure what is gained by calling all of these constructs "types of" blame. There are conceptual and psychological differences between blame and all of them (and among them), and our scientific theories and measurements must capture these differences. For example, on the basis of the four properties of blame we argued that blame is not just an emotion; even the best candidate emotion, anger, is a cousin, not a sibling to blame.

However, Sheikh and McNamara also mention *guilt* and *shame*, which seem closely related to blame—and for good reason, because a community that does not have to actively blame its members for every violation but can let some of the work be done by feelings of guilt will be all the more successful in keeping its members in line with its norms. Some may regard guilt as "internalized" blame. Nonetheless, in the process of guilt "becoming" (if it does) a first-person analog to blame, several transformations occur. Guilt arises much more often for unintentional violations, and immediately so, whereas it may take time to arise for intentional violations—because intentional violations are by definition in line with the person's goals at the time. Further, guilt serves both as self-regulation—trying to prevent the unintentional violation next time—and as social regulation—expressing guilt as an appeal to others to *reduce* their blame. The agent's expressed guilt may even stand in a hydraulic relationship[1] with the other's expressed blame: the less the offender expresses guilt, the more others blame; the more the offender expresses guilt, the less others blame (Ohtsubo & Watanabe, 2009; Scher & Darley, 1997). Finally, whereas people demand warrant for blame, it is unlikely they do for guilt. In fact, humans are capable of persistent guilt despite others' assurance that they absolutely do *not* blame the person. The classic case is survivor guilt, and a less extreme case would be a driver's guilt over killing a motorcyclist even though the latter ran a red light at 80 mph.

Shame, already differing from guilt, further differs from blame. For one, shame is character focused rather than event focused, as with blame or guilt (see Alicke, this issue). As a result, the community's devaluing of a person's trait may induce shame even without the person's own initial "detection" of a norm violation. The social-cognitive apparatus also seems underused with shame, as objects of shame can be physical traits, possessions, or social memberships. Moreover, the regulatory function of shame is murky, and some have even argued that shame is often dysfunctional (Tangney & Dearing, 2002).

We believe that pointing to similarities between blame and other phenomena is a valuable first step in grasping the network of interwoven moral phenomena. In fact, Voiklis, Cusimano, and Malle (2014) found that acts of moral criticism can be plotted in a similarity map that is akin to a network of family relation: containing blame's siblings (e.g., *reproach*), cousins (e.g., *condemn*), and neighbors (e.g., *punish*). But rather than treating neighbors like family, collapsing distal phenomena under the same overarching label, we believe the scientific task is to find the distinctions among these phenomena, allowing us to make fine-grained predictions that collapsed constructs do not permit.

Blaming How?

In the same spirit as differentiating blame from its family and neighbors, we would also like to differentiate actual "subtypes" of blame that some commentators highlighted or that we feel must be clarified: the difference between cognitive and social blame, between subjective and objective information processing in blame, and between blaming and providing warrant for blame.

[1]We thank Fiery Cushman and Julia Frankh for initial discussions about this topic.

Cognitive versus social blame. We proposed that cognitive blame as a private judgment and social blame as a public expression of blame are distinct but fundamentally entwined phenomena. We further suggested that social blame is the primary means by which people regulate others' behavior—that is, keep others in line with community norms. We strongly believe that moral psychology must study both forms of blame and especially their intriguing relationships, and we are particularly pleased that many commentators found this to be a worthwhile future direction for moral psychology.

DeScioli and Bokemper (this issue) help further characterize public blame by advancing the compelling idea that social blame signals willingness to coordinate with others. We take this to be coordination in two ways: the blamer's coordination with the offender, whom she offers an opportunity to repair his norm breach; and the blamer's coordination with other community members in the joint adherence to those norms.

Bauman and Mullen (this issue) help further characterize the function of private blame by suggesting that it helps reduce the social perceiver's risk of being exploited. By keenly detecting others' norm adherence and violations, perceivers learn to identify those who are trustworthy and those who are not and to selectively interact and form alliances with the trustworthy ones (Orbell & Dawes, 1993). Note, however, that merely *detecting* norm violations would deliver little diagnostic information. To gain information that is maximally diagnostic of the agent's future norm-relevant behavior, the perceiver must distinguish between negative events that the agent actually caused and those with which he was only associated, between those he brought about intentionally and those he caused unintentionally, and so on, for the remaining blame criteria. This illustrates again the importance of separating event-type moral judgments (something bad happened) and person-directed moral judgments (he deserves blame). The first may help avoid bad outcomes; the latter helps people strategically manage their future interactions with others.

Objective versus subjective processing of event information. Several commentators pointed to examples in which ideology or other person factors appeared to influence how people perceive norm-violating events, thus suggesting that people are not always objective in processing information about these events (Alicke, this issue; Nadler, this issue, citing Kahan, Hoffman, Braman, Evans, & Rachlinski, 2012). As masterfully documented in the social psychology classic "They Saw a Game" (Hastorf & Cantril, 1954), people with different attitudes frequently look at the same "objective situation" but detect different violations. We are in complete agreement here: People are, of course, not objective perceivers of reality, and all kinds of factors will influence how people conceptualize events, which features they attend to, and which norms they measure the event against. The Path Model is a psychological model—it describes how people subjectively conceptualize, represent, and process the information at hand, not how they process "objectively evaluated inputs" (Schein & Gray, this issue, p. 237). Against Alicke's characterization of our proposal, then, the Path Model is entirely compatible with "expectancy-driven processes in which people's prior knowledge of the actors involved, their social categories, and the type of event significantly influence the way *events are perceived and judged* [emphasis added]" (Alicke, this issue, p. 189).

Blame versus warrant for blame. When people blame an agent for a norm-violating event they are sometimes asked by others to provide warrant—"to offer grounds for why the agent deserves the attributed blame" (p. 149). Acceptable grounds, we argued, can refer to the event itself and the norm it violated but more importantly to the agent's causal involvement, intentionality, mental states, obligations, and capacities—the informational criteria we identified in the Path Model of Blame. This demand for warrant, we suggested, puts constraints on both the social expression of blame and the cognitive process of blame. It constrains the expression of blame because it establishes a norm: We ought not to blame people without being able to point to relevant causal and mental information. Moreover, it constrains the cognitive judgment of blame because people have to gather the kind of causal and mental information that can meet demands for warrant. The clear prediction is that when people expect to publicly declare their blame, they will carefully process the very information that counts as acceptable warrant for blame—which, if the Path Model is correct, are the criteria of causality, intentionality, reasons, and preventability.

Cushman (this issue, p. XX) raises the interesting question whether the social demand for warrant directly motivates or selects specific blame criteria. Here is what we believe such a selection process from the social to the cognitive level might look like:

1. A human community's need to make its members adhere to norms → selects for blame as a regulatory process.
2. The need to maximize voluntary norm adherence and to avert detrimental effects of haphazard, widespread blame → selects for targets of blame that are *real community threats* (exempting from blame those norm violations that serve the greater community good—i.e., "justified" violations) and only threats that *humans can*

control (exempting violations that even a willing community member cannot prevent).
3. The social-cognitive makeup of humans → selects for identification of the aforementioned blame targets by specific criteria: Actual threats are identified by inferring the agent's *reasons* and controllable threats are identified by counterfactual reasoning about agents' *capacity to prevent* future violations.
4. Demands for warrant of social blame → select for cognitive blame that meets these social-cognitive criteria.

If this analysis is correct, warrant itself does not select for criteria of blame, but it requires blamers to search for the criteria that maximize voluntary norm adherence.

We should note that the communicative act of providing warrant is not always a direct reflection of the private judgment; the social act of defending blame may be influenced by factors that are quite separate from the original blame judgment itself. We see this, for example, in Study 2 by Sood and Darley (2012), discussed by Nadler (this issue), where some participants had to warrant their blame in light of the "rule" that only harmful acts were blamable. This rule did not influence their blame judgments, but it did influence their warrants (claiming that they "saw" more harm), clearly dissociating blame from warrant for blame. False accusations are also cases of dissociation, where the accuser presents criterial information that would normally warrant blame even though the blamer does not believe this information to be true.

Applications of the Model: Victim Blame and Self-Blame

Cases of unquestioned blame that require theoretical explanation are victim blame and self-blame. We appreciate several commentators' encouragement to probe how well the Path Model applies to such important real-world cases as blame for victims of sexual assault. Although we hinted at the way our model would speak to these phenomena (p. XX), Niemi and Young's and Sheikh and McNamara's comments give us an opportunity to more directly apply the theory to victim blame, and Ames and Fiske suggested to do the same for self-blame.

Niemi and Young (this issue) pose the challenge to the theory this way: "If it is not a desire to blame the victim in service of sexism or the maintenance of purity and authority norms" (p. 232), what is victim blame? We would argue (very much in line with Sheikh and McNamara's analysis, p. XX) that people blame sexual victims when they hold certain assumptions about key criteria of blame—in particular, causality and preventability. These assumptions, which illustrate very well the role of "presets" (p. XX) operate in service of sexism and the defense of male dominance: They posit that women make *causal contributions* to rape, and they have *obligations* and *capacities* to prevent it. Thus, when rape occurs, in this view of the world, women are partially blameworthy. We can see these assumptions operate in the context of the horrific but deeply insightful events at Patrick Henry College. From Feldman (2014): "Responsibility falls disproportionately to women, who are taught to protect their 'purity' and to never 'tempt' their brothers in Christ to 'stumble' with immodest behavior" (p. 35). And therefore "the school puts the 'burden' on female students to ward off the male gaze" (p. 35). Thus, attractive dressing, being open, and all the normal flirting behaviors can later be interpreted as having causally instigated rape through temptation and as failing in the obligation to prevent men's sexual advances. Rape is, to a considerable extent, seen as the woman's fault because, "men only do bad things to impure women who have tempted them" (Feldman, 2014, p. 36).

This bundle of disconcerting presets has been collected systematically in rape myth scales. According to Burt (1980), rape myths are "prejudicial, stereotyped, or false beliefs about rape, rape victims, and rapists" (p. 217). And indeed, some of these myths formulate victims' causal contributions, obligations, and implied capacities to prevent. Consider these items from the Illinois Rape Myth Acceptance Scale (Payne, Lonsway, & Fitzgerald, 1999):

- When women go around wearing low-cut tops or short skirts, they're just asking for trouble.
- If a woman goes home with a man she doesn't know, it is her own fault if she is raped.
- If a woman is raped while she is drunk, she is at least somewhat responsible for letting things get out of control.

Recognizing the powerful impact of such preset assumptions on judgments of blame also helps explain why people can so vehemently disagree over the appropriateness of blame for a given event. Perceivers who share certain presets will come to one conclusion (e.g., the victim is partially to blame because she violated obligations of prevention), whereas perceivers who don't share those presets come to a very different conclusion (e.g., the victim has no obligation to prevent such events and deserves no blame whatsoever). It is also easy to see how two perceivers who disagree on criterial assumptions will regard the other's blame judgment as severely biased (Graham, Haidt, & Nosek, 2009).

And, as Sheikh and McNamara highlight, because these assumptions and consequent judgments are socially shared and recommunicated, they propagate within communities and become increasingly discrepant between communities.

In this way, blame is often constituted from a mix of stored representations (e.g., about obligations and general patterns of causality) and online processed information (e.g., about situation-specific preventability). We may then explain ideologically biased judgments of victim blame, not by postulating intrinsically biased online processing of event information, but by postulating previously adopted (often false) assumptions that operate as presets during judgment formation. People may well make reasonable efforts to process the event information at hand, but some people carry domain-specific presets that guide their information processing toward certain conclusions. We don't have to postulate a rampant "desire to blame"—which would lead to disastrous consequences in communities that exert strong social pressures for coordination, justice, and fairness (DeScioli & Bokemper, this issue)—but rather the influence of presets (sometimes correct, sometimes incorrect) and a certain miserliness of fresh information processing. In the *Motivated Blame* section of this Reply, we discuss further advantages of giving *presets* the explanatory burden of handling motivated reasoning.

This analysis of victim blaming applies similarly to a victim's self-blaming. We suggested in the previous section that guilt and blame share many properties but also differ in a few interesting ways. If self-blame equates to guilt, the same differences apply (though guilt may actually be less sensitive to warrant than self-blame is; Parkinson & Illingworth, 2009). In any case, the informational criteria that the Path Model specifies should apply to self-blame: Specifically, beliefs about causal contributions and preventability should increase self-blame, and these effects have been documented numerous times (Dalgleish, 2004). Further, self-blame, like victim blame, is influenced by community presets, such as Patrick Henry College's astounding maxims. Feldman (2014) again: "In the Christian world Claire had been brought up in, men only do bad things to impure women who have tempted them. She blamed herself. . ." (p. 36). Communities pay high costs when individuals excessively blame each other, but communities pay few costs when individuals excessively blame themselves. Consequently, whereas social norms regulate other-blame (especially through demands for warrant), there is little social regulation of self-blame and, perhaps at times, even encouragement by the dominant community members that the less dominant ones blame themselves.

The Power of the Event Concept

In our target article we tried to highlight some of the interesting aspects of the event detection phase, often overlooked in the previous literature. The commentators raised interesting additional aspects: the difference between single and multiple events (Goodwin; Nadler) and the appropriate theoretical treatment of unusual events—mental states, attempts, and dispositions (Alicke; Cushman; DeScioli & Bokemp; Goodwin).

Multiple Events

Several discussions of motivated reasoning touched on an ambiguity about how many events perceivers evaluate. Researchers often assume that when they provide participants with a single blame question, that blame question is answered in reference to a single event (let us call it the "target event," such as the oxygen tank explosion in one of Nadler & McDonnell's, 2012, cases). Often, however, researchers experimentally elicit a purported motivation to blame by telling participants something about some other event (e.g., that the protagonist, a football coach, had illegally administered the oxygen to his players). Given that this other event is itself a norm violation (and was intentional and unjustified), participants undoubtedly blame the protagonist for it. But they rarely have an opportunity to assign blame for this prior event; typically they are asked only about the target event (the explosion). One could therefore interpret participants' behavior as reflecting motivated blaming: Participants want to give more blame to the bad protagonist for the main event because the protagonist is a bad person. But one could also interpret this as a form of pragmatic responding (Adams & Steadman, 2004; Guglielmo & Malle, 2010): Participants give more blame to the bad protagonist for the totality of information they have learned about this person, and the only way they can do that is via the single question about blame for the target event.

Compound Events

Nadler (this issue) comments on what we originally called compound events (p. XX). She writes,

> Consider a driver who is speeding and kills a pedestrian as a result. The death was caused accidentally rather than intentionally. . . . But if the driver was a teenager who was speeding to try to show off for his friends in the car, we are likely to blame differently than if the driver was a father rushing his injured child to the hospital. (p. 227)

We fully agree, but not because "even for accidents we sometimes demand reasons"[2] (Nadler, this issue, p. 227) but because the perceiver is presented with a compound event—the teenager committed two violations. He both accidentally killed a pedestrian and was intentionally speeding for highly unjustified reasons. The father, by comparison, performed only one violation: accidentally killing the pedestrian, whereas his preceding action was justified and blameless. Nadler provides a similar analysis, calling such situations "a mixture" of intentional and unintentional events. We would say that the total situation consists of multiple norm-violating events, and blame is computed for each event, based on distinct information for the intentional and the unintentional event. (We used this approach to account for wayward-causation cases; p. XX).[3] Note that almost all cases of negligence and recklessness are compound events. There is usually (a) an unintended outcome (e.g., damage to the environment) and (b) an intentional action preceding the outcome (e.g., a CEO's decision to adopt a new program). Blame will vary depending on whether the action caused the outcome, whether the outcome was preventable by alternative actions, and whether the action had justified reasons. Thus, distinct information is recruited for each event, and overall blame should be sensitive to variations in facts about either event.

Unusual Events

Besides multiple and compound events, commentators also raised questions about how blame for unusual events should be handled, such as attempts (Cushman; Goodwin) and dispositions (Alicke; Goodwin; Sheikh & McNamara). We have applied our theory to norm-violating events that include outcomes, behaviors, mental states, and attempts and are aware of no data that contradict the model for those cases. We did, however, at one point claim (p. XX) that event evaluation does not require theory of mind capacities. As Cushman (this issue) rightly points out, this cannot be entirely true. When norm-violating events are mere outcomes (e.g., a dead body lying on the street), no social-cognitive work is needed to detect the norm violation. But when the event in question is a behavior (e.g., she yelled at him) or a mental state (e.g., he is planning to kill her), some minimal social-cognitive capacity must be involved, at least to recognize the event *as* a behavior or *as* a mental state. In any case, people do seem to blame people for mental states that have no outcomes whatsoever. In recent experiments, we found that people not only blamed fictitious others for thoughts, desires, or plans toward a harmful action *A* (e.g., selling drugs to teenagers) but their blame linearly increased from thinking about *A* to wanting to *A* to planning *A* (and was of course highest when *A* was actually performed). This finding also suggests that blame is motivated to a significant extent by the goal of preventing future violations, not just reacting to past harm, because in none of the mental-state events is there a negative outcome; nonetheless, the probability that harm will occur increases from thoughts to desires to plans, and so increases blame. This function of future prevention is perhaps a distinguishing feature of blame judgments, in contrast to wrongness judgments, which may have a more classifying rather than regulating purpose.

Because our theory focuses on blame for events (which are temporally delimited), we have no strong position on whether devaluing people for their character, incompetence, or other dispositions counts as blaming. At the conceptual level, we can see at least three properties of blame met by character disapproval (it is person directed, it relies on social cognition, and it demands some warrant). We are less certain about the property of social regulation. We proposed that second-person blame (unlike anger or punishment) invites or at least allows communication and possible repair of the fractured relationship between perpetrator and victim (or observer). By contrast, telling someone to their face "You are incompetent" or "You are a bad person" does not seem to invite communication or repair. But if we accept character at least as a target of cognitive blame, we must ask whether the core informational components of blame are taken into consideration. Many character judgments are themselves made on the basis of a person's reasons for his actions, so here the model fits. Likewise, dispositions such as incompetence are made primarily on the basis of failures to live up to one's obligations (though not necessarily moral ones). Perhaps the decisive question is whether people blame others for their dispositions only when they believe those dispositions to be malleable (i.e., "preventable"), because only then would the key function of blame—to regulate that which is blamed—gain any traction. The prediction would be that people don't blame someone for being schizophrenic (an unchangeable disposition), but they may blame someone for having a short temper or for being lazy (more changeable dispositions). Indeed, Weiner

[2]Like *blame*, the terms *reason* and *reasons* have multiple meanings in everyday language. Throughout the target article we used the terms to refer to the agent's mental states (typically beliefs and desires) in light of which and on the grounds of which he or she acted, following our theory of behavior explanations (Malle, Knobe, O'Laughlin, Pearce, & Nelson, 2000; Malle, 1999, 2004). See Malle (2011) for a detailed discussion of this theory and its supporting evidence.

[3]An interesting empirical question arises over whether the sum of people's blame is equal to the blame for each constituent event. We expect that, once the Path Model is formalized, a parameter of adjustment for compound events may be needed.

(1995) collected substantial evidence that variations in perceived preventability predict blame for various moral and nonmoral dispositions, such as cancer, AIDS, addiction, and obesity.

Motivated Blame: Toward a Convergence

One major theme in the commentaries is motivated cognition. Apparently we took a controversial position by swinging the emphasis from the countless flaws of human judgment to the sophisticated information processing that underlies judgments of blame. We proposed that the default information processing en route to blame follows a canonical (though not inevitable) path of specific conceptual categories. But we emphasized that this information processing 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" (p. 152). So we hardly "banish the role of motivation in blame judgments" (Nadler, this issue, p. 222) or suggest that blame is "immune" to the possibility of motivated reasoning (Cushman, this issue, p. X; Nadler, this issue, p. XX). The Path Model not only allows for motivated cognition, affective biases, and the like, but it offers a specific analysis of *how* these motivational and affective processes influence blame. Indeed, Bauman and Mullen (this issue) pointed out that motivated reasoning is perfectly compatible with our model, Niemi and Young (this issue) anticipated how the model would account for motivated victim blame, and Ames and Fiske (this issue) agree that the Path Model can provide predictions about the ways in which "motivation and cognition might interact to affect moral judgment" (p. 193).

We are optimistic that continued discussion and debate will lead to a further convergence of perspectives and to increasingly powerful explanatory models. On this path to convergence we would like to discuss two phenomena that we believe fail to illustrate motivated blaming (outcome bias and early desire to blame) and then explicate how the Path Model characterizes interactions between motivation, affect, and blame.

Outcome Bias

Outcomes matter. They define what norms were violated, and they provide evidence about various blame criteria. But responsiveness to outcomes is not the same as outcome "bias." There is room for all kinds of inaccuracies when people attend to and interpret outcome information; but surely people are not *mistaken* when taking into account outcome information en route to blame. When Cushman (this issue) states that "people tend to judge an action more harshly if it happens to lead to a more harmful outcome" (p. 203), he is not describing a bias but a natural ingredient of blame. To demonstrate actual bias, some kind of distinction needs to be established between outcome information that is valid and outcome information that is not valid (Malle et al., this issue, p. X). Even Nadler (this issue), who does not entirely agree with our view on motivated blame, is pessimistic about the possibility of establishing such a normative distinction (p. XX).

Alicke (this issue) suggests that the Path Model underestimates "the role of outcome information on evaluations of various blame criteria" (p. 188). To the contrary, we argue that knowing about the "outcome"—the context, behaviors, and persons involved—is *required* for processing information about the various criteria. How else should the perceiver learn about what happened, who was involved, and whether the agent knew one thing or wanted another? There is room for selective and motivated processing of this information. But even with imperfect processing, the Path Model delineates what kinds of information people search for and consider *before* they arrive at a blame judgment.

Desire to Blame, Early

According to one dominant picture of motivated reasoning, people have a general desire to blame, from the start. When they encounter an opportunity to blame, they have a "preferred conclusion" (that the person in question be blamed) and then adjust their subsequent information processing to fit the conclusion (Alicke, this issue; Nadler, this issue; see also Ditto, Pizarro, & Tannenbaum, 2009; Sood & Darley, 2012). Critically, in this hypothesis the preferred conclusion is formed *before* information processing occurs, and it *directly* produces a blame judgment: "Sometimes perceivers begin with a sense of blame, and work backwards through the information provided to find the harm or other circumstance that justifies blame" (Nadler, this issue, p. XX). For Alicke, too, blame criteria are "often calibrated or adjusted a posteriori rather than assessed a priori; that is, after rather than before knowledge about a harmful outcome has been received" (p. 188).

In such a picture we would have to explain how people come to their early "preferred conclusions" without information processing. This seems mysterious, even paradoxical. To compute just a sense of blame (e.g., for certain persons), considerable information must be acquired—at a minimum, that the event was a norm violation (so the moral system gets activated); whether an agent was involved (so the desire to blame gets triggered); who the agent was (so a group bias is triggered); and, if the desire to blame is graded, what the agent specifically did.

Moreover, a desire-to-blame model cannot account for the very results that seem to motivate it. How do we explain why some people (due to ideology or other individual differences) selectively blame some groups and not others? For example, why do some people select to blame the victim rather than the perpetrator?[4] Why do Republicans blame social welfare politicians and Democrats blame corporations for unemployment? Both groups have a general desire to blame and engage, it is argued, in motivated reasoning; why would they arrive at entirely different conclusions?

The Power of Presets

As we suggested in the *Applications of the Model* section, such divergent conclusions due to ideology or other perceiver characteristics can be best explained by the operation of preset beliefs about blame-relevant information—for Republicans and Democrats, for example, about what causes unemployment, who has obligations to prevent it, and by which means. These beliefs, when plugged into a generally stable information-processing apparatus, will predictably lead to divergent judgments of blame.

The concept of presets in the Path Model explains several other phenomena. First, it explains how ambiguity can permit motivated reasoning. When information is ambiguous or absent, presets are easily plugged in; when the information is unambiguous or undeniable, it overrides presets.

Second, presets are helpful in accounting for the impact of prototypical roles such as villain versus victim (Schein & Gray, this issue). Past encounters may build a picture of the person as one prototype but, more important, these past encounters shape agent-specific knowledge structures: the person's past behaviors, mental states, and typical situations in which they find themselves. Such knowledge structures are quickly activated when the perceiver sees that person act or suffer again, and these presets, along with the specific information at hand, will guide the resulting blame judgment.

Third, Sheikh and McNamara (this issue, p. XX) made us realize that the picture of presets also instills some optimism about the possibility of reforming distorted judgments in strongly ideological groups. If what distorts people's processing of rape, poverty, and other victimizations were their general desire to blame, then we would have no hope for change, because "that's just how people are." But if the processing is not inherently biased but influenced by specific presets (e.g., assumptions as documented in rape myth scales), one could target such incorrect assumptions, educate people about the facts, and thereby change blame judgments without having to change their basic mechanism of blame processing.

Blame as a Working Hypothesis

Alicke (this issue) puts it well: "Blame is a working hypothesis" (p. 258), but a working hypothesis, we would add, that is invoked by information processing. For example, before people blamed O. J. Simpson for killing his wife and lover, they obviously had to *learn* that the wife and her lover had been brutally murdered and that Simpson was fleeing from the police. There was no early blame *before* they learned this information; there could have been no judgment "a priori"—that is, "before knowledge about a harmful outcome has been received" (Alicke, this issue, p. 188).

Similarly, when Mark's hiking partner lets go of a tree branch that smacks Mark in the face (Alicke, this issue, p. X), blame does not emerge ex nihilo; several pieces of information are preset or available to Mark at the very moment of observing the event: Hiking partners have an obligation to prevent things like that; his partner had no intention to smack him. A working hypothesis of blame can be computed quickly from this information. The key test is whether the working hypothesis is modulated by criterial information that emerges a moment later: Blame should decrease if the partner is genuinely surprised that Mark was behind him (unintentional, low preventability → low blame), and blame should increase if the partner laughs gleefully (intentional, mean reason → high blame). Such tests were provided by Monroe and Malle's (2014) studies, in which participants formed an initial blame hypothesis on limited information (about agent, behavior, and some likelihood of intentionality) and, after receiving further information, they updated their hypothesis. People were highly sensitive to the newly incoming information, increasing or decreasing their blame in systematic ways.

The notion of a working hypothesis also fits with Spellman and Gilbert's (this issue) suggestion to treat the information processing toward blame as dynamic and iterative, in which later acquired information can update earlier acquired information, and in which counterfactual reasoning can inform both capacity considerations ("He could have . . .") and considerations of actual causal contributions. Dynamic updating, however, does not mean reverse processing order: for example, it makes little sense to explore the agent's reasons before an intentionality hypothesis is formed; and once intentionality is established, counterfactual reasoning about preventability is not relevant.

[4]See Krahé's (1988) data, which show that *most people* give a 0% rating on the measure of a victim's responsibility for being raped. This is puzzling under the postulate of a general desire to blame.

Blame as working hypothesis should also address Goodwin's (this issue) concern that the Path Model might not "account for findings showing that blame judgments may bias later judgments of causation or intention" (p. 217). We argue that blame *judgments* already have to be made on the basis of causal-mental information, but these judgments may bias later *reports* of causality, intentionality, and so on (as in Nadler & McDonnell, 2012, Study 2), especially if the reports are made under demand for warrant (Sood & Darley, 2011). Unfortunately, assessing the temporal-causal separation of judgments and reports is a serious challenge, as is the temporal-causal separation of one kind of judgment from another. Measuring causality after blame and showing their statistical dependence does not demonstrate such a causal-temporal sequence (as Spellman and Gilbert, this issue, highlight). In particular, when all the causal-mental information is contained in the stimulus, people do not even have the option of making a blame judgment without considering the information right in front of them. Motivational biases would be more convincingly demonstrated by showing that people *ignore* hypothesis-inconsistent information even if the information is right in front of them.

However, despite our concerns about claims of motivated blaming, we believe there is genuine convergence between the Path Model of Blame and the various perspectives that highlight motivated reasoning. And this is what we close with.

Paths of Convergence

The first convergence is that the Path Model accommodates, as Goodwin (this issue) rightly demands, the influence of motivation on the *processing of* causal and mental information. Specifically, we suggested that such influence operates via "microprocesses" at each information component (p. XX): via *Concept activation* (e.g., a hurtful remark activates the intentionality concept); *Information acquisition* (e.g., inferential hyperactivity, retrieval of presets); and *Value setting* (lowering evidence thresholds). To study these kinds of *CIV* mechanisms will require new studies with fine-grained measurement approaches.

The second path of convergence is that the Path Model of Blame includes a candidate for early motivational or affective force—just not early *blame*. Observing a norm-violating event often leads to an *evaluation,* sometimes a strongly affective one, which could have a detrimental impact on subsequent processing. Researchers must measure such early evaluation before additional information is presented in order to demonstrate biasing effects on subsequent information processing (ideally by pinpointing which microprocesses are affected and how).

A third path of convergence is implicit in some commentaries and our target article: that whatever motivational biases cognitive blame may suffer from, once blame is socially expressed it will come under the scrutiny of other people's countervailing biases, demands for warrant, and fact checking. These mechanisms will ultimately have to succeed at fine-tuning the validity and fairness of moral criticism so it can achieve its function of social regulation and cooperation. The study of social blame promises to be a challenging and exciting direction of future research.

Coda

We greatly appreciate the critical and creative commentaries that our theory elicited, which have sharpened our own understanding of blame. The theory also inspired diverse evaluations (sometimes within the same commentary), ranging from ones that modesty forbids us to recite to the suggestion that the theory "fails to deliver a compelling new set of principles." For our part, we are most certainly motivated and biased evaluators of our theory; nonetheless, we feel that several of its elements are not "retrograde" but do represent new and synthesizing ideas:

- that social processes of regulation and demand for warrant constrain cognitive blame;
- that intentionality bifurcates information processing and predicts which kinds of information people will seek and be sensitive to;
- that preset information can speed up, even bias the judgment process, and that the concept of presets offers an account of individual differences in blame judgments; and
- that motivated reasoning operates through a set of information criteria and their constitutive microprocesses (concept activation, information acquisition, value setting).

At first we were disappointed to learn that our theory of blame "lacks the "hook" that great music and successful theories share" (Alicke, this issue, p. 191). But at least two of us don't particularly care for pop music, for which a "hook" is so desirable. We hope instead that scientific theories can be more like ~~classical music~~, representing a complex and partially unpredictable world that is nonetheless structured by fundamental regularities—regularities of the kind that scientists of morality are converging to understand.

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

References

Adams, F., & Steadman, A. (2004). Intentional action in ordinary language: core concept or pragmatic understanding? *Analysis*, *64*, 173–181. doi:10.1111/j.1467-8284.2004.00480.x

Burt, M. R. (1980). Cultural myths and supports for rape. *Journal of Personality and Social Psychology*, *38*, 217–230. doi:10.1037/0022-3514.38.2.217

Dalgleish, T. (2004). What might not have been: An investigation of the nature of counterfactual thinking in survivors of trauma. *Psychological Medicine*, *34*, 1215–1225. doi:10.1017/S003329170400193X

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Moral judgment and decision making.* (Vol. 50, pp. 307–338). San Diego, CA: Elsevier Academic.

Feldman, K. (2014, March 3). The price of purity: Sexual assault at God's Harvard. *The New Republic*, *244*, 32–41.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*, 1029–1046. doi:10.1037/a0015141

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

Hastorf, A. H., & Cantril, H. (1954). They saw a game: A case study. *The Journal of Abnormal and Social Psychology*, *49*, 129–134. doi:10.1037/h0057880

Kahan, D. M., Hoffman, D. A., Braman, D., Evans, D., & Rachlinski, J. J. (2012). "They saw a protest": Cognitive illiberalism and the speech-conduct distinction. *Stanford Law Review*, *64*, 851–906.

Kahn, P. H., Jr., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., Shen, S., … Severson, R. L. (2012). Do people hold a humanoid robot morally accountable for the harm it causes? *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 33–40). New York, NY: ACM. doi:10.1145/2157689.2157696

Kahn, P. H., Jr., Ruckert, J. H., Kanda, T., Ishiguro, H., Reichert, A., Gary, H., & Shen, S. (2010). Psychological intimacy with robots? Using interaction patterns to uncover depth of relation. *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 123–124). Piscataway, NJ: IEEE. doi:10.1109/HRI.2010.5453235

Krahé, B. (1988). Victim and observer characteristics as determinants of responsibility attributions to victims of rape. *Journal of Applied Social Psychology*, *18*, 50–58. doi:10.1111/j.1559-1816.1988.tb00004.x

Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, *3*, 23–48. doi:10.1207/s15327957pspr0301_2

Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: MIT Press.

Malle, B. F. (2011). Time to give up the dogmas of attribution: A new theory of behavior explanation. In M. P. Zanna & J. M. Olson (Eds.), *Advances of experimental social psychology* (Vol. 44, pp. 297–352). San Diego, CA: Academic Press.

Malle, B. F., Knobe, J., O'Laughlin, M. J., Pearce, G. E., & Nelson, S. E. (2000). Conceptual structure and social functions of behavior explanations: Beyond person-situation attributions. *Journal of Personality and Social Psychology*, *79*, 309–326. doi:10.1037/0022-3514.79.3.309

Monroe, A. E., Dillon, K. D., & Malle, B. F. (in press). Bringing free will down to earth: People's psychological concept of free will and its role in moral judgment. *Consciousness and Cognition*.

Monroe, A. E., & Malle, B. F. (2014). *Moral updating*. Unpublished manuscript, Brown University, Providence, Rhode Island.

Nadler, J., & McDonnell, M.-H. (2012). Moral character, motive, and the psychology of blame. *Cornell Law Review*, *97*, 255–304.

Ohtsubo, Y., & Watanabe, E. (2009). Do sincere apologies need to be costly? Test of a costly signaling model of apology. *Evolution and Human Behavior*, *30*, 114–123. doi:10.1016/j.evolhumbehav.2008.09.004

Orbell, J. M., & Dawes, R. M. (1993). Social welfare, cooperators' advantage, and the option of not playing the game. *American Sociological Review*, *58*, 787–800. doi:10.2307/2095951

Parkinson, B., & Illingworth, S. (2009). Guilt in response to blame from others. *Cognition and Emotion*, *23*, 1589–1614. doi:10.1080/02699930802591594

Payne, D. L., Lonsway, K. A., & Fitzgerald, L. F. (1999). Rape myth acceptance: Exploration of its structure and its measurement using the Illinois Rape Myth Acceptance Scale. *Journal of Research in Personality*, *33*, 27–68. doi:10.1006/jrpe.1998.2238

Scher, S. J., & Darley, J. M. (1997). How effective are the things people say to apologize? Effects of the realization of the apology speech act. *Journal of Psycholinguistic Research*, *26*, 127–140. doi:10.1023/A:1025068306386

Scheutz, M. (2012). The inherent dangers of unidirectional emotional bonds between humans and social robots. In P. Lin, G. Bekey, & K. Abney (Eds.), *Anthology on robo-ethics* (pp. 205–221). Cambridge, MA: MIT Press.

Sood, A. M., & Darley, J. M. (2012). The plasticity of harm in the service of criminalization goals. *California Law Review*, *100*, 1313–1358.

Tangney, J. P., & Dearing, R. L. (2002). *Shame and guilt*. New York, NY: Guilford.

Voiklis, J., Cusimano, C., & Malle, B. F. (2014, July). *A social-conceptual map of moral criticism.* Presented at the Proceedings of the 36th Annual Meeting of the Cognitive Science Society, Quebec City, Canada.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York, NY: Guilford.