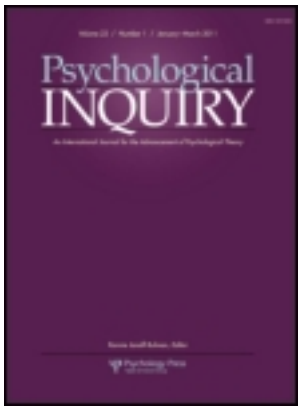

Psychological Inquiry: An International Journal for the Advancement of Psychological Theory
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hpli20

The Scope of Blame

Fiery Cushman[a]
[a] Department of Cognitive, Linguistic, & Psychological Sciences, Brown University, Providence, Rhode Island
Published online: 20 May 2014.

To cite this article: Fiery Cushman (2014) The Scope of Blame, Psychological Inquiry: An International Journal for the Advancement of Psychological Theory, 25:2, 201-205, DOI: 10.1080/1047840X.2014.904692

To link to this article: http://dx.doi.org/10.1080/1047840X.2014.904692

Psychological Inquiry, 25: 201–205, 2014

ISSN: 1047-840X print / 1532-7965 online
DOI: 10.1080/1047840X.2014.904692

Routledge
Taylor & Francis Group

The Scope of Blame

Fiery Cushman
Department of Cognitive, Linguistic, & Psychological Sciences, Brown University, Providence, Rhode Island

Malle, Guglielmo, and Monroe (this issue) do three great services to the field of moral psychology. First, they define a discrete kind of moral evaluation—blame—and characterize the distinctive cognitive architecture that sets it apart from judgments of wrongness, moral character and so forth. Second, they articulate the component elements of an extended, deliberative judgment process and thus offer a welcome contrast to a common fetishizing of gut intuitions in recent research on moral judgment. Third, they situate these psychological mechanisms within a social context and, in doing so, acknowledge that moral concepts are as defined by their interpersonal functions as by their individual representations. Any one of these accomplishments could easily have supported an article of its own.

And perhaps they should have. The Persian flaw of Malle and colleagues' otherwise expertly woven work is to insist that these three portraits—a discrete kind of evaluation, an extended judgment process, and a social interaction—in fact depict the same object from different points of view. Of course, it is easy to appreciate the motive for this proposed unification. The authors' three portraits are surely linked, and it would be a much graver mistake to analyze any of them in utter isolation from the others (as others have done). Still, the three concepts of blame that Malle and colleagues discuss are not identical. Rather, they are joined in the way that a vow, a wedding, and a marriage are joined. Although these objects surely cannot be considered in isolation of each other, neither can they be conflated. They operate on different time scales, and they encompass phenomena of vastly different scope. Distinguishing between three distinct concepts of blame can help to resolve a few of the lurking tensions that strain the coherence of Malle and colleagues' work.

A Discrete Kind of Moral Evaluation

The heart of Malle and colleagues' theory is the path model of blame. Consistent with a careful conceptual analysis as well as their own prior empirical research (Monroe, 2012; Monroe & Malle, 2014), it suggests a specific sequence of cognitive operations that transforms the perception of an event into a kind of moral evaluation. The model proposes that first a norm-violating event is perceived, next causal responsibility for the event is established, next the mental states of the causally responsible agent are assessed, and finally either justification or preventability is assessed, depending on the intent of the agent.

This model is both familiar and distinctive. It is familiar in the sense that its basic contours are shared with a long tradition of attributional models. These models include variants by Heider (1958), Shaver (1985), Weiner (1995), and others. And it is distinctive in the sense that all of these models capture a specific kind of moral evaluation, not "the" general process of moral judgment writ large. This particular kind of moral evaluation yields judgments of blame, of course, but also judgments of responsibility, liability, and punishment.

The defining feature of this category of moral evaluation is that it binds a person to an event via both causal responsibility and intent. Malle and colleagues place more emphasis on the "person" as the target of blame, but it is clear that the event is essential to their theory as well. In fact, it is the very genesis of blame attribution: One begins with an event that demands blame and then proceeds to identify and evaluate the responsible individual. In other words, blame reads like a detective novel. We begin with a body on the ground, the first question was ask is "Whodunnit?" and as soon as we've answered that question, we want to know why they did it. As Malle and colleagues (this issue) note, this stands in contrast to other categories of moral judgment, such as judgments of "wrongness." The object of a wrongness judgment is the action that a person performs, not their relation to an event. Wrongness judgments begin with that action and then proceed to consider the mental states that gave rise to that action. Consider people who string a wire across the sidewalk in order to trip old ladies: The action they perform is wrong, but they will be blamed for the outcome.

Malle and colleagues first highlight this distinction between punishment and wrongness judgments but then proceed to subtly undermine it. This occurs as they attempt to broaden the scope of what a theory blame should encompass. Consider, for instance, their discussion of attempted harms. An attempted harm involves an action but no outcome: For instance, you put what you believe is rat poison in a friend's food,

but really it is sugar. Should the path model apply to such cases?

At first blush it seems that it should not. Malle and colleagues (this issue) are quite explicit in stating that "moral event detection does not require theory of mind capacities" (p. 152). Rather, the events comprising a norm violation are supposed to be defined in terms of either a harmful outcome (e.g., a black eye) or a bad action (e.g., a fist swung at a face). There is certainly no norm violation in putting sugar in a friend's food, either in terms of the outcome or in terms of the physical action. Surely it is only the agent's mental states—her false belief that the sugar is poison—that could support to a negative evaluation. Thus, the natural conclusion is that the concept of blame simply does not apply to attempted harms. There is no body on the ground; no dunnit to who. This approach has been taken by past theorists. For instance, Darley and Shultz (1990, p. 531) wrote, "Judgments of moral responsibility presuppose those of causation. If the protagonist is judged not to have caused the harm, then there is no need to consider whether he is morally responsible for it."

Instead, Malle and colleagues (this issue) pursue a more expansive concept of blame. They accommodate attempts by arguing that the relevant "event" is simply the agent's own action or intention—the malevolent sugaring. They write,

> 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. (p. 153)

Later they elaborate on a similar theme:

> 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. (p. 168)

This approach succeeds at forcing the judgment of attempted harms into the path model, but at what cost? It manifestly contradicts the earlier claim that "moral event detection does not require theory of mind capacities" (p. 152), it weakens the concept of "event" to encompass a concept defined precisely by its noneventness (omission), and it makes hay of the very useful distinction between blame and wrongness. After all, the distinguishing feature of wrongness judgments is that it targets behaviors rather than the link between events and persons. This distinction is clearly lost when Malle and colleagues introduce the possibility that "the event is a behavior" (p. 153). Indeed, later on they explicitly consider conflating the concepts of wrongness and blame entirely: "Is wrongness a judgment sui generis or is it equivalent to a blame judgment of norm-violating actions?" (p. 159). An eager "yes" hangs on the lips of this rhetorical question, but to give in to "yes" is to give up on a precise and useful distinction. Malle and colleagues should have the courage of their convictions and affirm the their own conceptual differentiation between blame and wrongness, embrace their own model of blame as a category of evaluation that is uniquely triggered by true events, and therefore concede that it is not applied in any straightforward manner to attempted harms. This is not, of course, to say that attempted harms are not subject to moral evaluation. Rather, it is to say that that relevant moral evaluation is not usefully described as one of "blame"—at least, not the category of blame that the path model so elegantly describes.

An Extended Judgment Process

The concept of blame judgments as a desecrate kind has had tremendous influence over a long history of attribution research. One of Malle and colleagues' (this issue) most important contributions to this literature is to show how blame judgments are modified by a sequence of additional considerations, including preventability and justification. Their analysis of the distinction between preventability and controllability is particularly careful and convincing. Yet is the extended judgment process envisioned by Malle and colleagues in the late stages of their path model (preventability and justification) really identical to the discrete category of "blame" judgment that motivates the early stages of the model (assigning responsibility for an event to an agent)?

The authors give a strong hint that the answer must be "no" when they consider whether blame attribution is best characterized as an automatic or controlled process:

> 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. (p. 151)

This all-encompassing approach would be peculiar if Malle and colleagues were restricting their analysis to a single, well-defined moral evaluation—in other words, to a concept that picks out a discrete psychological process, such as we considered above.

That processs might be automatic or it might be controlled, but at least would be *something*. Rather, it reveals a second and quite distinct sense of "blame": A judgment process that unfolds over a long time scale and draws upon many subsidiary mechanisms. The process is defined more by the kinds of information that it ultimately will incorporate, by its function, and by its social context than by the discrete psychological mechanisms that may happen to support it on one occasion or another.

The concept of an extended process of moral judgment is sorely missing from current research in moral psychology and is likely to receive an especially warm welcome from philosophers. A common philosophical critique of current research in "moral judgment" is its single-minded focus of rapid, automatic gut intuitions. Perhaps unsurprisingly, many philosophers hold that worthwhile moral judgments are instead the product of much reasoning and deliberation. Malle and colleagues' second concept of blame has much in common with the philosophers', and much to offer the field of psychological research. If the experience of an automatic moral intuition is a common part of the human experience, surely the feeling of mulling over every dimension of a difficult moral dilemma is as well (Paxton, Ungar, & Greene, 2012; Pizarro & Bloom, 2003).

Perhaps as a consequence of their laudable vision of an extended process deliberative reasoning, Malle and colleagues take extraordinary pains to defend blame against nearly every accusation of bias and motivated reasoning that has been offered in the literature. Of course, there are good reasons to question the wisdom of their position on a priori grounds. If blame turned out to be immune to biases and motivated reasoning it would belong to a very rare category of psychological processes indeed. But there are also much more focused reasons to question its wisdom, and these hinge on the distinction between blame as a discrete category of moral judgment versus blame as an extended judgment process.

Much evidence suggests that moral judgments are susceptible to an outcome bias (e.g., Berg-Cross, 1975; Cushman, 2008; Cushman, Dreber, Wang, & Costa, 2009; Gino, Moore, & Bazerman, 2009; Gino, Shu, & Bazerman, 2010; Mazzocco, Alicke, & Davis, 2004). Specifically, people tend to judge an action more harshly if it happens to lead to a more harmful outcome compared with a less harmful outcome or no harm at all. Malle and colleagues (this issue) attribute a rational basis to this apparent bias: It arises, they claim, from the valid inference that harmful outcomes are a product of malicious intentions (or, at the very least, negligence). But there is good evidence that this explanation is not adequate; moreover, it obscures one of the chief advantages of considering blame as a discrete category of moral evaluation. When a person's intentions are crossed with the outcomes they bring about in an factorial design, judgments of blame and punishment show a strong outcome effect (about 20% of variance explained), whereas judgments of moral wrongness show a weak outcome effect (about 2% of variance explained; Cushman, 2008). In unpublished data, it was found that judgments of moral character are similarly weakly influenced by outcome effects. In other words, a strong outcome bias is one of the distinctive features that sets blame judgments apart as a distinct kind. Malle and colleagues' path model actually provides a very natural explanation for this distinctive feature: The basic function of a blame judgment is to link a person to a harmful event, and the detection of a harmful event comprises the initial starting point for a blame judgment. In addition, the discrepancy between blame/punishment judgments and wrongness/character judgments provides strong evidence against the hypothesis that outcome bias is exclusively due to the rational inference to culpable mental states. After all, there is no obvious reason why this inference would be more valid when a person judges blame, as compared to judging moral wrongness or character.

Finally, the "rational inference" model does not easily explain why people tend to assign less blame to attempted harms than fully completed harms. After all, both involve equally malicious intent. Malle and colleagues (this issue) acknowledge this point, offering that "the constitutive actions of [an attempted harm, such as robbing a bank] usually violate fewer or weaker norms than the constitutive actions of actually robbing the bank (the latter involving far more manifest damage)" (p. 168). Yet, if "manifest damage" occupies a central role in the process of blame attribution, and if it modifies the degree of blame assigned even holding constant a wrongdoer's intent, then we are back to the original concept of outcome bias that Malle and colleagues were at pains to deny.

This nettlesome contradiction can be avoided by differentiating between a restricted and automatic process of blame attribution that is strongly influenced by outcomes (independent of intent) and a more expansive and deliberative process of moral judgment that tends to override outcome biases in favor of intent-based moral judgment—one that is consistent with the normative standard of many philosophers and of Malle and colleagues as well. This division fits comfortably with recent evidence that outcome biases are enhanced under cognitive load (Buon, Jacob, Loissel, & Dupoux, 2013), as well as with the developmental trajectory of intent-based moral judgment (Cushman, Sheketoff, Wharton, & Carey, 2013; but see Hamlin, 2013; Hamlin, Ullman, Tenenbaum, Goodman, & Baker, 2013).

With this distinction in hand, we can also squarely confront one of the most interesting novel questions posed by Malle and colleagues' framework. Are the assessments of preventability and justification—the twin horns of the path model's bifurcation—core features of the automatic and desecrate process of blame attribution, or instead are they invoked only during the more extended and deliberative phase of moral evaluation? This is a key area for further research that follows directly from Malle and colleagues' model, if only we grant ourselves the latitude to make strong commitments to the roles of automaticity and control in moral judgment and to differentiate the narrow and broad senses of blame.

A Social Act

Finally, Malle and colleagues (this issue) invoke the concept of blame as a social act. There are few issues both so crucial to the study of moral judgment and so underappreciated in current research. Here, Malle and colleagues make a laudable foray into a promising new frontier.

Their approach reflects a long-standing debate in philosophy that concerns the nature of concepts. This debate asks whether we should think of concepts as fully constituted by the mental states that instantiate them, such that they would exist unchanged in a brain in a vat (the so-called "internalist" position), or whether instead we should think of them as depending crucially upon a connection to external features of the world, such as the referent of the concept or its functional role in a social context (the so-called "externalist" position). Current research on moral judgment tends to implicitly adopt an internalist position. These studies attempt to map out the representations and computations that mediate between a stimulus input and a behavioral response—precisely the kind of experiment that could be run on a brain in a vat, if only such a thing were available on undergraduate campuses and online labor marketplaces. In contrast, Malle and colleagues' urge us instead toward an externalist perspective. They argue that the very concept of blame depends upon its social function: The role it plays in shaping others' behaviors, signaling a person's preferences, and providing warrant for an attitude such as anger.

Like many philosophical debates, the battle between internalism and externalism seems to pose a false choice. Surely there is utility in characterizing blame both narrowly (as a psychological mechanism that mediates between stimulus and response) and broadly (as a kind of social behavior that fulfills an important functional role). Moreover, there is a clear relation between these levels of analysis. On one hand, the social function and context of blame can explain why the psychological process takes the form it does (whether through biological evolution, cultural evolution, or learning). On the other hand, the form of the psychological process will necessary constrain the range of social behaviors and outcomes that can be expressed. For instance, it has been argued that the process of assigning liability in the law is influenced not only by the functional and demands of a legal system but also by the psychological mechanisms that generate our moral intuitions (Mikhail, 2009).

Characterizing the interaction between these levels of analysis is an exciting new frontier in moral psychology precisely because it is meaningful to differentiate between them. (After all, without differentiation there can be no interaction.) A particularly intriguing question is whether the social demand for warrant directly motivates distinct elements of the path model at a proximate level. That is, do people consider an agent's justification for an act (for instance) because they are directly motivated by the need to justify blame to a social community?

Conclusion

Malle and colleagues' article defines the state of the art in blame attribution research. Their contribution is remarkable both for the focused nuance of its conceptual analyses and for the wide scope of the phenomena it attempts to capture. Yet these virtues sometimes work at cross-purposes. It is possible to achieve clear focus in capturing a discrete kind of moral evaluation, and in capturing an extended judgment process, and in capturing a socially embedded behavior. Yet it may not be possible to maintain that focus while attempting to unify all three. In the few places where these conceptions rub up against each other the clarity of Malle and colleagues' argument shows frayed edges; their treatment of attempted harms and outcome bias are two salient examples. Still, these points hardly detract from the substantial accomplishment of their important work.

Note

Address correspondence to Fiery Cushman, CPLS Department, Box 1821, Brown University, 190 Thayer Street, Providence, RI 02912. E-mail: fiery_cushman@brown.edu

References

Berg-Cross, L. (1975). Intentionality, degree of damage, and moral judgments. *Child Development*, *46*, 970–974.

Buon, M., Jacob, P., Loissel, E., & Dupoux, E. (2013). A non-mentalistic cause-based heuristic in human social evaluations. *Cognition*, *126*, 149–155.

Cushman, F. A. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*, 353–380.

Cushman, F. A., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a 'trembling hand' game. *PLOS One*, *4*(8), e6699. doi:10.6610.1371/journal.pone.0006699

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition*, *127*, 6–21.

Darley, J. M., & Shultz, T. R. (1990). Moral rules—Their content and acquisition. *Annual Review of Psychology*, *41*, 525–556.

Gino, F., Moore, D., & Bazerman, M. (2009, April 8). *No harm, no foul: The outcome bias in ethical judgments* (Harvard Business School NOM Working Paper No. 08-080). Available at SSRN: http://ssrn.com/abstract=1099464

Gino, F., Shu, L., & Bazerman, M. (2010). Nameless + harmless = blameless: When seemingly irrelevant factors influence judgment of (un) ethical behavior. *Organizational Behavior and Human Decision Processes*, *111*, 93–101.

Hamlin, J. K. (2013). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition*, *128*, 451–474.

Hamlin, J. K., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: experiments in preverbal infants and a computational model. *Developmental Science*, *16*, 209–226.

Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley.

Malle, B. F., & Holbrook, J. (2012). Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality and Social Psychology*, *102*, 661–684. doi:10.1037/a0026790

Mazzocco, P., Alicke, M., & Davis, T. (2004). On the robustness of outcome bias: No constraint by prior culpability. *Basic and Applied Social Psychology*, *26*, 131–146.

Mikhail, J. (2009). Moral grammar and intuitive jurisprudence: A formal model of unconscious moral and legal knowledge. *Psychology of Learning and Motivation*, *50*, 27–100.

Monroe, A. E. (2012). *Moral updating*. Unpublished dissertation. Brown University, Providence, RI.

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

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*, 163–177.

Pizarro, D. A., & Bloom, P. (2003). The intelligence of the moral intuitions: A comment on Haidt (2001). *Psychological Review*, *110*, 193–196.

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness.* New York, NY: Springer-Verlag.

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