Mindless, Harmless, and Blameworthy

David A. Pizarro
Department of Psychology, Cornell University, Ithaca, New York

David Tannenbaum
Anderson School of Management, University of California, Los Angeles, Los Angeles, California

Eric Uhlmann
Management and Human Resources Department, HEC Paris—School of Management, Paris, France

There is a compelling simplicity to the theoretical approach to moral judgment proposed by Gray, Young, and Waytz (this issue; henceforth GYW). On their approach, all that is needed to account for the large body of empirical findings on moral judgment is a description of the prototypical moral encounter—a moral agent who brings harm to a moral patient. This is what psychological theorizing ought to look like: explaining the observed complexity of a phenomenon by appealing to more basic, general, psychological mechanisms.

However, the simplicity of the dyadic approach outlined by GYW may not be sufficient to account for several recently documented aspects of moral judgment. Namely, that there are a number of situations in which neither agency nor harm (as typically defined) appear necessary for the ascription of moral responsibility and blame. For instance, in our own work we have documented cases in which individuals judge a transgression to be morally wrong despite a clear absence of harm, as well as cases in which individuals are deemed to be blameworthy despite their lack of agency (e.g., Inbar, Pizarro, & Cushman, 2012; Tannenbaum, Uhlmann, & Diermier, 2011). Elsewhere, we have argued that these results are best explained by a fundamental feature of moral evaluation—the motivation to assess an individual's underlying moral character (Pizarro & Tannenbaum, 2011). On this person-centered account, the question at the heart of moral evaluation is, Is this a *good* person or a *bad* person? This approach, which emphasizes the goal of evaluating persons over time, stands in contrast to many extant theories of moral judgment that emphasize moral evaluations made for singular acts (often inspired by deontological and consequentialist normative theories; Bayles, 1982) and that characterize moral judgment as seeking to answer a different set of questions such as, Was a moral rule broken? Was any harm caused? and Did the person have an intention to act harmfully?

To be sure, the Morality as Mind Perception account (hereafter MMP) defended by GYW reliably accommodates many intuitions about how individuals arrive at judgments of moral blame and responsibility, such as the intuition that individuals who intended harm are more blameworthy than those who did not, and that harmless acts often do not even qualify for moral evaluation. But the person-centered approach, we have argued, is better suited to account for recent findings demonstrating that even actions that are harmless or mindless (i.e., not performed with obvious agency) are nonetheless judged to be blameworthy.

To its credit, the MMP account goes beyond current theories of moral judgment in its acknowledgment of character as a central feature of moral evaluation (albeit in a supporting role). However it still relies fundamentally on the claim that morality is, at its essence, about the dyadic relationship between a moral agent and a moral patient. For instance, GYW contend that "all moral transgressions are fundamentally understood as agency plus experienced suffering—i.e., interpersonal harm—even ostensibly harmless acts such as purity violations" (p. XX). It is possible, as GYW argue, that even in the absence of explicit harm, moral transgressions (such as eating an already-dead dog; Haidt, 2001; Haidt, Koller, & Dias, 1993) may be implicitly viewed as causing symbolic harm and suffering. Yet from a person-centered perspective, it is not necessary to stretch the definition of harm in this fashion in order to account for such acts. Because even harmless acts can be informative of an individual's underlying character, such acts are important candidates for moral evaluation. It is to these harmless (and in some cases mindless) acts that we turn.

Harmless but Informative

Harmful acts are often informative of moral character, and it is intuitively appealing to think that the magnitude of harm ought to directly relate to the moral depravity of an individual. For example, it seems obvious that a man who murders his wife is a morally worse person than a man who physically abuses his wife, and this man is worse than a man who doesn't abuse his wife at all. Yet there are examples where there is no clear monotonic relationship between harm

and character. For instance, Tannenbaum et al. (2011, Experiments 1a and 1b) asked participants to evaluate two individuals: a man who, upon learning that his girlfriend has been unfaithful, reacts by beating his girlfriend's cat, or a man who (in the same situation) reacts by beating his girlfriend. Whereas participants viewed the latter action (beating a girlfriend) as more harmful and morally worse, they viewed the man who beat his girlfriend's cat as having a worse moral character (i.e., possessing more negative traits in the domain of moral character).

Yet in this example both individuals (the cat beater and the girlfriend beater) clearly brought about harm to a patient, and both were judged as morally bad individuals. Perhaps the MMP account does not require a strict monotonic relationship between harm and judgments of moral character—only that an agent cause harm to a patient. However, there are also examples where agents are not causally responsible for harm, or where no harm occurred, yet these behaviors still seem informative about an agent's underlying moral character. Such acts are often drawn from the domain of moral purity—something that the MMP approach attempts to account for—but not always. For instance, Inbar et al. (2012) recently demonstrated that gaining financially from a harmful event (e.g., by betting that a natural disaster will occur within the next year) is judged as a blameworthy act, and individuals who engage in these actions are seen as possessing bad character. In these cases, it is clear that there is no way an individual would have been capable of having caused the harm. It was sufficient that he gained from it.

Mindless but Informative

Just like harm, goal-directed mental states such as beliefs, desires, and intentions—the critical features of agency—can be informative of moral character (indeed, one sense of the term "character" is a set of stable intentions to act morally or immorally). As evidence of the diagnostic value of intentions for assessing character, one need only compare an individual who causes harm unintentionally (e.g., by tripping over a dog) to an individual who causes the same harm intentionally (e.g., by kicking a dog). It is of little surprise that intentions are weighted heavily in both judgments of moral responsibility and in assessments of character.

But intentions are not always the best guide to an individual's character. This is true, in part, because people may be motivated to signal good intentions even when those intentions are not the true source of their behavior. This motivation to engage in impression management by broadcasting one's intentions has been noted by attribution theorists for decades (Jones & Pittman, 1982; Jones & Whortman, 1973). For instance, a restaurant patron may leave an oversized tip not because he is a generous soul but merely to impress his date. For this reason people sometimes view unbroadcasted behavior as particularly revealing of character. For instance, consider a high school student who tells his parents and friends that he is attending parties every Friday night, when in actuality he is spending those nights purchasing and distributing food to the homeless. If the high schooler's secret activities were to be unexpectedly exposed—say his friends drive past and witness him feeding the homeless—this would presumably be more meaningful than if the teenager had announced his activities at school the day before. The difference between the two seems clear: In the first instance the student likely acted out of morally pure intentions, whereas in the second instance there could be a number of reasons for the behavior (the student could be acting from pure intentions or simply attempting to impress his teachers enough to write him a good letter of recommendation). This process of reasoning backward from an effect to its underlying cause is inherent to judgments of character; we determine the probability that an individual possesses a particular character attribute given that they performed a certain behavior. In the example of the altruistic high schooler, the informativeness of the unbroadcasted behavior comes from the fact that it rules out an alternative cause for the behavior (e.g., that it was driven by impression management). Not only is this a normatively reasonable strategy, as ruling out alternative causes can only increase our confidence in the target cause (e.g., behavior X was caused by character attribute Y), people also appear quite attuned to these considerations when assessing social behavior (Fernbach, Darlow, & Sloman, 2010, 2011).

In a similar vein, mindless behaviors (such as unconscious or automatic behaviors that lack the presence of intentions) can be informative about moral character in that they exclude a variety of conscious motives that might otherwise present plausible alternative explanations for a behavior (such as the sorts of impression management motives just described). Because they provide a more "pure" source of information about an underlying trait, a mindless behavior can amplify moral evaluations in a manner that is not predicted by either extant theories of moral responsibility, nor, as we understand it, by the MMP. A number of fairly mundane examples illustrate this point: Imagine a scenario in which a spouse becomes extremely drunk (so much so that she loses consciousness and "blacks out"), and while on the cab ride home she tells her husband that she never loved him, that he's a wimp and a pushover, and that he could never make her happy. It is likely that most individuals would see this as blameworthy behavior and as extremely revealing (both about her character and about her attitude toward her husband). Yet in an important sense the spouse did not meet the criteria for agency during this episode.

Legal judgments about negligence follow a similar logic. In U.S. and English common law, negligence is sometimes determined not by using "subjectivist" standards (i.e., appealing to an agent's state of mind at the time of the act) but rather according to "objectivist" standards (i.e., what should we expect of a reasonable person in that situation?; Duff, 1990, chap. 7). Although the law recognizes that an agent may not have mindfully produced a harm (e.g., when a parent forgets that his child has been left out in the cold), individuals are deemed blameworthy because of the counterfactual (e.g., any reasonable person would not forget about his children). Presumably, our reaction to this sort of event is evidence that we have gleaned valuable information—that the person is lacking the "right" set of values and priorities.

Similarly, quick, impulsive decisions (which are less mindful than slow, deliberative decisions) can be viewed as more blameworthy precisely because they are thought to be revealing about an agent's values (Critcher, Helzer, Tannenbaum, & Pizarro, 2012; Critcher, Inbar, & Pizarro, under review). For instance, Critcher et al. (2012) provided subjects with descriptions of agents engaging in morally questionable behaviors. When the agent committed the act without hesitation and with little emotion (so-called "rash" behaviors), participants viewed the activity as morally worse than when he engaged in the same behavior after careful consideration. In these cases, participants viewed rash decisions as more revealing about the individual's underlying preferences.

Character and Anticipated Future Harm

It appears as if agency and harm are not always necessary for moral judgment in the manner predicted by the MMP account. Yet there may be common ground between the MMP and the person-centered account of moral judgment we have been defending here. It may be that moral transgressions are ultimately linked to concerns about material harm, but in some cases the concerns are not about the harm that has occurred but about the probability that a person will engage in future harm. Specifically, negative character inferences based on seemingly harmless transgressions may be geared toward anticipating and preventing future harms by that person. It seems a reasonable prediction that an individual who engages in seemingly harmless acts such as sex with dead animals (Haidt, 2001, Haidt et al., 1993) may be an unstable, antisocial, and dangerous person who likely cannot be trusted to do things like cooperate in vital group endeavors and refrain from harming ingroup members. Consistent with this general idea, an individual who engages in certain putatively harmless behaviors may nonetheless be seen as morally flawed and may receive the same moral evaluations given to agents who engage in clearly harmful behaviors. For instance, Tannenbaum et al. demonstrated that a CEO candidate who requests a harmless but offensive perk (i.e., an expensive marble table with an engraving of his face) is also viewed as more likely to cause serious damage to the company in the future (Tannenbaum et al., 2011, Experiment 2). What remains unclear at this point, however, is whether perceptions of future harm are the primary drivers of moral evaluations of character.

In sum, the basic motivation to diagnose an agent's underlying character is an account of moral judgment that, in our view, does not require broadening the concepts of "agency" and "harm" to accommodate the available data. This is not to say that the MMP account is without value to understanding person-centered moral judgment; several theoretical frameworks have begun to elucidate the ways in which mental state inference and conceptions of agency are linked to broader assessments of persons (Malle, 2004; Reeder, 2009), and the framework provided by GYW helps to enrich this picture. That said, we believe that the relationships between agency, harm, and character are complex and there is still plenty of conceptual work to be done. We believe that the real value of the MMP account to understanding character still awaits, and will be determined by the degree it can uniquely articulate the relationship between agency and character while also generating fresh insights for future research.

Note

Address correspondence to David A. Pizarro, Cornell University, 224 Uris Hall, Ithaca, NY 1853-7601. E-mail: dap54@cornell.edu

References

Bayles, M. (1982). Character, purpose and criminal responsibility. *Law and Philosophy, 1*, 5–20.

Critcher, C. R., Helzer, E. G., Tannenbaum, D., & Pizarro, D. A. (2012). *When doing good isn't good enough: Moral judges credit principled actions, not actions that promote principle*. Unpublished manuscript, University of California, Berkeley, CA.

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2012). *How quick decisions illuminate moral character*. Unpublished manuscript, University of California, Berkeley, CA.

Duff, A. (1990). *Intention, agency and criminal liability: Philosophy of action and the criminal law*. Oxford, UK: Blackwell.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science, 21*, 329–336.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General, 140*, 168–185.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

3

310 Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65*, 613–628.

Inbar, Y., Pizarro, D. A., & Cushman, F. (2012). Benefiting from misfortune: when harmless actions are judged to be morally
315 blameworthy. *Personality and Social Psychology Bulletin, 38*, 52–62.

Jones, E. E., & Pittman T. S. (1982). Toward a general theory of strategic self-presentation. In J. Suls (Ed.), *Psychological perspectives on the self* (Vol. 1, pp. 231–262). Hillsdale, NJ: Erl-
320 baum.

Jones, E. E., & Whortman, C. (1973). *Ingratiation: An attributional approach*. Morristown, NJ: General Learning Press.

Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction.* Cambridge, MA: MIT Press. 325

Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). Washington, DC: American Psychological 330 Association.

Reeder, G. D. (2009). Mindreading: Judgments about intentionality and motives in dispositional inference. *Psychological Inquiry, 20*, 1–18.

Tannenbaum, D., Uhlmann, E. L., & Diermeier, D. (2011). Moral 335 signals, public outrage, and immaterial harms. *Journal of Experimental Social Psychology, 47*, 1249–1254.

4