

Blaming Badly

Mark D. Alicke

Department of Psychology, Ohio University, Athens, OH, 45701, USA,
e-mail: alicke@ohio.edu

Abstract

Knobe and Burra show that people will ascribe intentionality to outcomes that an actor had no intention to effect. I describe this ostensibly contradictory effect in terms of outcome bias – the tendency for people to blame actors for the outcomes of their actions, independent of their causal and intentional role in the event. Whereas intention refers to an actor's pre-behavioral desires, intentionality characterizes an event that includes both behavior and outcome. Thus, while people recognize that the actor did not desire for the outcome to occur, they ascribe intentionality as a means of registering disapproval for what ultimately *did* occur. I describe these findings in terms of my culpable control (Alicke, 2000) model of blame.

Keywords

Blame, intention, culpable control

Moral philosophers, legal theorists and psychologists who study moral judgment are remarkably agreed in prescribing how to blame people. A blameworthy act occurs when an actor intentionally, negligently or recklessly causes foreseen, or foreseeable, harmful consequences without any compelling mitigating or extenuating circumstances. This simple formulation conveniently forestalls intricacies about how to construe concepts such as will, causation, foresight and mitigation, but, putting that aside for the moment, it seems fair to say that blame “professionals” share compatible conceptions of how to make fair and rational blame ascriptions.

Blame “amateurs”, on the other hand, sometimes stretch these rational prescriptions for blame beyond the breaking point. Numerous psychological studies on blame and responsibility, as well as the perplexing outcomes of high-profile criminal and civil trials, demonstrate that everyday blamers are capable of violating virtually every rational prescription that moral philosophers, legal scholars, and rational decision theorists hold dear.

Intention and Intentionality

The papers in the special issue of *Journal of Cognition and Culture* (vol. 6/1–2) on folk conceptions of mind, agency and morality focus, among other things, on one way in which ordinary blame or moral judgments potentially diverge from rational prescriptions, namely, by calling outcomes “intentional” that they believe actors had no “intention” to effect. Knobe and Burra’s paper leaves little doubt that such attributions occur. In one of their vignettes, a company director decides to initiate a program that will increase profits but that will also harm the environment. About 87% of the research participants who read this story agreed that the director harmed the environment intentionally, but only 29% agreed that he had an intention to do so. On the other hand, when this same vignette was framed positively, such that the new program could inadvertently help the environment, only 20% characterized the director’s action as intentional and hardly anyone believed that he intended to be helpful.

The explanation that Knobe and Burra favor for this asymmetry involves the moral status of the director’s behavior. I agree with this assumption, but I also concur with Knobe and Burra that various unanalyzed complexities lurk behind this seemingly straightforward claim. Granting, as I do, that the moral (or what I would call evaluative) status of behavior and its outcomes, as well as the mental states that precede and accompany them, can influence people’s judgments of intention (and virtually every other aspect of a harmful act), the obvious questions are: (i) why this occurs in Knobe and Burra’s scenario and (ii) how significant these findings are for the psychology of how people judge intent and intentionality, and for philosophical construals of these constructs.

Let me begin by saying that I think the more noteworthy aspect of Knobe and Burra’s findings is the difference between intent and intentionality imputations in the harmful outcome scenario. I consider the asymmetries the authors found between the harming and the helping scenarios to be interesting but somewhat less significant. Here’s why. Knobe and Burra’s findings show that the collateral effect of the director’s decision on the environment constitutes a basis for blame when it is harmful, but provides scant basis for praise when it is beneficial (as reflected in judgments of intentionality). This suggests that people have a much lower threshold for attaching collateral effects to an actor’s moral account when the effects are negative *versus* positive. This seems to me like an eminently sensible decision strategy. The director’s avowed goal in both the harmful and helpful scenarios is to make a profit. In the harmful scenario, he pursues this goal with the knowledge that collateral harm will surely occur. His actions are akin to the legal offense of recklessness.

Reckless behavior occurs when people are willing to risk collateral damage to achieve a desired end. As in Knobe and Burra's example, reckless individuals foresee the possibility of the collateral damage but they do not desire it. Reckless people can be held criminally responsible or civilly liable depending on the severity of the risk they incurred, the probability that the risky consequences would materialize, and the amount of damage that occurred. Punishing reckless behavior is, therefore, dignified in Anglo-American law, well-entrenched in everyday moral intuition and perhaps even helpful in discouraging future transgressions. Because reckless people are dangerous, they are blamed not only for undesired outcomes, but even for outcomes whose probability is much lower than that of the collateral damage in Knobe and Burra's examples, which was virtually certain. But whereas blaming reckless people has obvious instrumental value, there is little to be gained in praising inadvertent do-gooders. After all, it is not as though we can facilitate their future beneficence by extolling them for fashioning consequences that they could not have cared less about.

Of course, Knobe and Burra's actual findings pertained to the tendency to call undesired outcomes "intentional", rather than "blameworthy". I have assumed, along with other commentators on their paper, that the willingness to say that collateral harm occurred intentionally, and the refusal to use the same label for collateral benefits, represents propensities to blame, and to refrain from praising, respectively. And this, I think, is the really noteworthy aspect of their findings, and one that deserves further attention.

To begin with, Malle and Knobe's research (1997) shows that the ordinary, or "folk" concept of intention is multi-faceted. The main elements of intent in the present context are desire and foresight. Although the contributors to this volume tend to describe desire and foresight as things that people either possess or lack, I assume that they would allow for gradations of these mental components. The CEO in Knobe and Burra's scenario has virtual certainty that the collateral effect of damaging the environment will occur, and so his degree of foresight is very high. And while he has no specific desire to damage the environment and probably wouldn't do so willy-nilly, he is more than happy to let the damage occur if it satisfies his financial motive. The desire component, therefore, is quite low.

Now, intentionality is a very different beast. Whereas desire and foresight are mental states that precede and guide behavior, intentionality is a post hoc characterization of an action and its consequences. As such, intentionality judgments are qualitatively different from judgments of intent because they include knowledge of how the event actually turned out. Put another way, whereas intent judgments in Knobe and Burra's example pertain to what the

CEO was willing to let happen, intentionality judgments refer to what actually did happen. Intentionality judgments are, therefore, susceptible to outcome bias, which involves judging an outcome in terms its consequences, independent of the decision-maker's intentions and of the causal process by which they were generated (Baron and Hershey, 1988; Alicke and Davis, 1989). Outcomes exert an enormous influence on blame and responsibility judgments. In our initial studies on this topic (Alicke and Davis, 1989), participants read stories about a homeowner who heard noises in an upstairs bedroom, got a licensed gun and went to investigate, encountered a man going through his daughter's dresser, and shot the presumed intruder when he turned around quickly. In the positive scenario, it turned out that the victim was indeed a burglar with a long criminal record. In the negative scenario, the victim was the boyfriend of the homeowner's daughter who was picking up some clothes for her. Participants blamed the homeowner more when he shot the innocent victim than when he shot the dangerous criminal.

In more recent studies (e.g., Mazzocco and Alicke, 2005), we have shown that the tendency to blame the homeowner more for harming an innocent victim obtains even when there had been a rash of burglaries in the area and his neighbor had been robbed at gunpoint two days earlier; when he had never used a gun but went to check on his neighbor's child because his neighbor was out of town and an escaped criminal had just killed a homeowner in the neighborhood; and when he first tried to call the police, and instead of using a gun, simply called after the intruder to stop as he was descending the stairs, which scared the intruder and caused him to fall down the stairs, hit his head, and suffer brain damage. The outcome bias effect is, therefore, extremely robust; in fact, it is more challenging to eliminate the effect than it is to demonstrate it.

So when research participants are asked to determine whether the CEO in Knobe and Burra's scenario intended to harm the environment, they typically say no, because the question they are answering is whether he desired this consequence. When they are asked whether he harmed the environment intentionally, however, they are answering a question that includes the outcome that he did, in fact, harm the environment. Imputations of intentionality in the absence of assumptions about intent are, according to this view, due to the incorporation of outcome information into intentionality judgments.

Culpable Control

The foregoing explanation invites the question of why parallel effects do not occur in the helpful scenario. In other words, if people use negative out-

come information to punish the CEO by ascribing intentionality, why don't they do the same thing to praise him when a positive outcome occurs? To account for this attributional asymmetry, I rely on a brief description of my culpable control model (Alicke, 1992, 2000), especially the blame validation assumption.

When people assess blame, they estimate the degree to which an actor controlled (or in the case of incomplete offenses, would have controlled) the harmful outcomes. The aspects of control that laypeople consider entail the same criteria that moral theorists prescribe for assessing blame (i.e., desire, foresight, foreseeability, causal influence, absence of coercion, lack of justification). However, each evidential item is also assessed spontaneously along a positive-negative evaluative dimension. This spontaneous evaluation assumption is based on the well-established finding that evaluation is the most fundamental component of human judgment (Osgood *et al.*, 1957), and on extensive research showing that positive and negative evaluations are automatically evoked in relevant contexts (Fazio *et al.*, 1986; Bargh and Chartrand, 1999), such as in the present one of assessing harmful behavior.

Whereas Knobe and Burra's CEO who has no qualms about harming the environment is likely to elicit strong negative spontaneous reactions, the fortuitously helpful CEO is unlikely to inspire positive ones. According to the blame validation assumption of the culpable control model, once strong negative reactions have been evoked, people view the relevant evidence in a way that justifies their desire to blame the source of those reactions. In an early series of studies (Alicke, 1992), we argued that causal judgments can be swayed by blame validation processing. In the initial experiment, we presented participants with scenarios in which a student had a car accident while he was speeding home from school. In one version, the student's reason for speeding was to hide a vial of cocaine from his parents that he had carelessly left on a table; in a second version, his motive was to hide his parents' anniversary present. The student was judged to have had a significantly greater causal role in the car accident when his motive was to hide a vial cocaine as opposed to an anniversary present. In other studies, we have found that other components of blame such as intent and foresight (Alicke *et al.*, 1990) are subject to similar blame-validation influences.

The phrase "culpable control" reflects the assumption that the desire to hold someone culpable slants perceptions of the control evidence in a direction that validates blame. However, the degree to which control assessments can be skewed is limited by reality constraints. Even potent negative evaluations will fail to enhance blame without a sufficient evidential basis. For example, a virulent racist will probably decline to blame an African-American for an

offense in which he had absolutely no causal involvement. In Knobe and Burra's example, the CEO's callous disregard for the environment elicits, according to the culpable control perspective, a blame-validation mode of evidence evaluation, but the evidence for his having an intention (i.e., a desire) to harm the environment is so remote as to discourage this inference, even if people are inclined to make it. On the other hand, judgments of intentionality, which are based on judgments about the event itself, including the outcome, rather than on the CEO's intentions, are more amenable to spontaneous evaluation influences.

Some Loose Ends

The papers in the special issue also address two other main questions. The first question concerns the process by which people use positive and negative moral evaluations as a basis for assessing blame criteria such as intention. The second question is whether the fact that people make moral evaluations for events that are completely determined by natural laws supports the philosophical position of compatibilism, that is, that people can be held morally responsible for their actions even if those actions are completely determined by past and present causal conditions. Let me begin with the second question, because unless I am missing something, the answer to this seems straightforward.

When it comes to prescribing how moral judgments should be made, lay intuition is an invaluable resource. Moral and legal theorists can and should debate the merits of things like holding people criminally responsible for unintended outcomes (as in negligence), for grading completed attempts more severely than incomplete ones, and for punishing people who suffer from various mental disorders and diminished capacities. Decisions about these issues cannot avoid ordinary intuitions about fairness and justice. The Anglo-American legal system has evolved over many years to accommodate such intuitions. So, while I would not endorse using public opinion surveys to make and revise laws, I think it is advisable for the legal system to attend to lay intuition.

On the other hand, perennial debates about free will, and the specific problem about the relation between free will and moral responsibility, seem largely beyond the auspices of ordinary intuition. We would hardly take the fact that the vast majority of the United States population believes that we have been visited by aliens as a reason for revising our beliefs about alien visitations. The fears, superstitions and cognitive tendencies that these beliefs reflect are fascinating in many ways, but they are not very important in helping us to under-

stand interplanetary communication. The perplexing departures of ordinary moral judgment from rational prescriptions fall into the same category: they are necessary to understanding how folks go about making such determinations, but they have little to contribute to the logical analysis of compatibilism-incompatibilism with regard to moral responsibility.

At last, a couple of closing comments on the tendency to predicate blame assessments on moral values rather than strictly on the rational criteria that underlie moral philosophy, jurisprudence and rational decision theory. In his paper, Malle discusses the work of Adams and Steadman (2004) who argue that the folk concept of intentionality represents peoples' desires to blame the CEO in Knobe's scenarios. This is obviously very close to the blame validation assumption in the culpable control model. Whereas Adams and Steadman describe this tendency as a pragmatic one, stemming from the desire to express disapproval of the CEO's behavior, the culpable control model views this tendency as even more basic, deriving from the automatic positive and negative evaluations that accompany perceptions of value-relevant events. Spontaneous evaluations are often, perhaps even usually, inaccessible to conscious awareness. For example, I doubt that many of the participants in the accident scenario I described earlier (Alicke, 1992) were aware that their heightened perceptions of the actor's causal role in the car accident were due to the fact that his goal was to hide a vial of cocaine from his parents. This lack of awareness can partially explain the low to moderate correlations that Malle reports between intent and blame. In our research, we have tended to obtain stronger relationships between criteria such as intent, causation, negligence and foresight on the one hand, and blame on the other, and so the specific experimental contexts in which these variables are studied probably contribute to the magnitude of this relationship, as Malle notes in his paper.

I should also note that spontaneous evaluations are not identical to emotional reactions. Emotions can certainly heighten one's positive and negative evaluations to information about an event, but the tendency to evaluate experiences and events automatically can be "cool" as well as "hot." Thus, the neuropsychological evidence reported by Young *et al.* in their paper, which suggests that emotion (measured in seven brain-damaged patients) is not necessary to explain the heightened intentionality judgments reported by Knobe and Burra, while quite interesting, does not speak to the role of spontaneous evaluations in such judgments.

One last thing. In accord with Malle's observations, I would encourage investigators who are working on these problems to provide more detailed, varied, and realistic stories to research participants. The situations described in these papers include very short, hypothetical stories, which run the risk of

encouraging participants to treat them as games, thereby raising questions as to whether these response represent their true attitudes. I have often obtained different results when using more involved stimulus materials, and when leading participants to believe that they are judging actual instances of wrongdoing that have, for example, been reported in the newspaper, or that will soon be adjudicated.

References

- Alicke, M.D. (1992). Culpable causation. *Journal of Personality and Social Psychology* 63, 368-378.
- (2000). Culpable control and the psychology of blame. *Psychological Bulletin* 126, 556-574.
- Alicke, M.D. and Davis, T.L. (1990). Capacity responsibility in social evaluation. *Personality and Social Psychology Bulletin* 16, 465-474.
- (1989). The role of a posteriori victim information in judgments of blame and sanction. *Journal of Experimental Social Psychology* 25, 362-377.
- Alicke, M.D., Weigold, M.F. and Rogers, S.L. (1990). Inferring intentions and responsibility from motives and outcomes: Evidential and extra-evidential judgments. *Social Cognition* 8, 286-305.
- Adams, F. and Steadman, A. (2004). Intentional action in ordinary language: Core concept or pragmatic understanding? *Analysis* 64, 173-181.
- Bargh, J.A. and Chartrand, T.L. (1999). The unbearable automaticity of being. *American Psychology* 54, 462-479.
- Baron, J. and Hershey, J.C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology* 54, 569-579.
- Fazio, R.H., Sanbonmatsu, D.M., Powell, M.C. and Kardes, F.R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology* 50, 229-238.
- Malle, B.F. and Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology* 33, 101-121.
- Osgood, E.E., Suci, G.J. and Tannenbaum, P.H. (1957). *The measurement of meaning*. University of Illinois Press, Urbana, IL.

Copyright of Journal of Cognition & Culture is the property of Brill Academic Publishers and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.