Cognition 117 (2010) 139–150

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Enough skill to kill: Intentionality judgments and the moral valence of action

Steve Guglielmo *, Bertram F. Malle

Brown University, Department of Cognitive, Linguistic, and Psychological Sciences, 89 Waterman St., Providence, RI 02912, United States

ARTICLE INFO

Article history:
Received 22 November 2008
Revised 2 July 2010
Accepted 7 August 2010

Keywords:
Moral judgment
Blame
Intentionality
Social cognition
Bias
Causal deviance

ABSTRACT

Extant models of moral judgment assume that an action's intentionality precedes assignments of blame. Knobe (2003b) challenged this fundamental order and proposed instead that the badness or blameworthiness of an action directs (and thus unduly biases) people's intentionality judgments. His and other researchers' studies suggested that blameworthy actions are considered intentional even when the agent lacks skill (e.g., killing somebody with a lucky shot) whereas equivalent neutral actions are not (e.g., luckily hitting a bull's-eye). The present five studies offer an alternative account of these provocative findings. We suggest that people see the morally significant action examined in previous studies (killing) as accomplished by a basic action (pressing the trigger) for which an unskilled agent still has sufficient skill. Studies 1 through 3 show that when this basic action is performed unskillfully or is absent, people are far less likely to view the killing as intentional, demonstrating that intentionality judgments, even about immoral actions, are guided by skill information. Studies 4 and 5 further show that a neutral action such as hitting the bull's-eye is more difficult than killing and that difficult actions are less often judged intentional. When difficulty is held constant, people's intentionality judgments are fully responsive to skill information regardless of moral valence. The present studies thus speak against the hypothesis of a moral evaluation bias in intentionality judgments and instead document people's sensitivity to subtle features of human action.

1. Introduction

Some observations about moral judgment are uncontroversial. Foreseeability has been widely observed as a requirement for folk responsibility judgments—people normally do not consider someone responsible for outcomes that the person was unable to anticipate (Shaver, 1985). Equally significant, Hamilton (1978) recognized the role of obligation, as people are blamed only for negative outcomes that they were obligated to prevent. Finally, Weiner (1995) observed that such obligation is meaningful only if the outcome was controllable by the person—that is, if he or she could have intentionally prevented it. Thus, blame for negative events arises when the person *should have* and *could have* prevented it (Malle, Moses, & Baldwin, 2001).

These findings highlight people's rational, rule-following assignments of responsibility and blame (cf. Nichols & Mallon, 2006). Admittedly, things are not always so clear-cut. Alicke (2000) and others have shown that extraneous variables can alter moral judgment beyond rational rules (Schnall, Haidt, Clore, & Jordan, 2008). But there is little doubt that such rules exist and operate in many instances.

One of the most important rules governing the assignment of blame is that intentional moral transgressions—when a person intentionally performs an immoral action—amplify blame (Cushman, 2008; Heider, 1958; Ohtsubo, 2007; Shaver, 1985). Unintentional harmful

* Corresponding author. Address: Brown University, Department of Cognitive, Linguistic, and Psychological Sciences, 89 Waterman St., Providence, RI 02912, USA. Tel.: +1 401 863 2727.
E-mail address: Steve_Guglielmo@Brown.edu (S. Guglielmo).
URL: http://research.clps.brown.edu/mbq/guglielmo/ (S. Guglielmo).

0010-0277/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2010.08.002

behavior may elicit blame if the agent could have and should have prevented the harm (Weiner, 1995). But actually having a desire and intention to bring about harm, and exerting effort to realize this intention, is the worst offense in any social community. Excuses sometimes avert blame for unintentional harm; only justifications can possibly avert blame for intentional harm (Tedeschi & Reiss, 1981), and except for rare cases in which a compelling reason for the harmful act is available (e.g., the dentist hurting her patient), full blame applies.

Judgments of intentionality, thus, serve as a central input to judgments of blame. Schematically, the social perceiver takes two judgment steps, in order (Guglielmo, Monroe, & Malle, 2009):

1. Determine: Is the negative event intentional?
2. If Yes → examine the actual intention or goal; then assign proper blame.
   If No → examine obligation and ability to prevent (foreseeability and controllability); then assign proper blame.

Knobe (2003a, 2003b) challenged this fundamental order and proposed instead that the badness or blameworthiness of an action can influence people's intentionality judgments. In particular, Knobe argued, the same behavior that is seen as unintentional when performed without moral implication (e.g., shooting and hitting a bull's-eye) may in fact be seen as intentional when performed with moral implication (e.g., shooting and killing another person). If true, such a pattern would cast serious doubt on rational models of both blame assignment and intentionality judgments. People would not, as traditionally believed, assess intentionality to designate blame but would instead assess blame to designate intentionality.

The fault of the traditional account of blame, following Knobe's argument, lies in its assumption about how intentionality judgments are made. Both philosophical theories (e.g., Mele, 1992; Mele & Sverdlik, 1996; Searle, 1983) and psychological theories of intentionality (Malle & Knobe, 1997) subscribe to the valence-neutral model sketched in Fig. 1. According to this model, for positive, negative, and neutral behaviors alike, people process five information components that all have to be present for a behavior to be considered intentional. If even one component is missing, the behavior does not count as intentional.

Knobe claims that this five-component model is correct only for neutral actions but does not hold for negatively valenced actions (we will return to the question of morally positive actions later in this article). Knobe's evidence for his claim falls into two sets. The first set challenges the necessity of the *intention* component for judgments of intentionality. According to the standard intentionality model, a behavior is judged as intentional only if the agent actually intended to perform that behavior (Adams, 1986; Malle & Knobe, 1997). In contrast, Knobe (2003a) provided data suggesting that people consider a known but *unintended* side effect intentional if that effect is negatively valenced (e.g., harming the environment, risking the lives of soldiers, decreasing sales). This effect has been replicated numerous times (Cushman & Mele, 2008; McCann, 2005; Nadelhoffer, 2006a; Nichols & Ulatowski, 2007), and we refer to this set of findings as the "side-effect challenge."

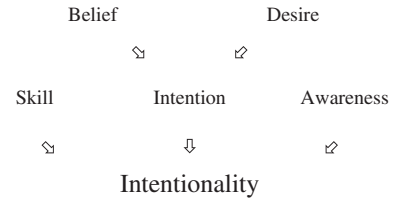

Fig. 1. A model of the folk concept of intentionality.

The second set of findings questions the necessity of the *skill* component for judgments of intentionality. According to the standard intentionality model, people judge a behavior as intentional only if the agent has reliable ability or skill to produce that behavior (Malle & Knobe, 1997; Mele & Moser, 1994; Thompson, Armstrong, & Thomas, 1998). Knobe (2003b) showed that an agent's unskilled neutral action (e.g., a lucky shot to win a contest) is not viewed as intentional but an equivalent unskilled *immoral* action (e.g., a lucky shot to kill someone) is very much seen as intentional. Once more, other researchers have replicated this effect (Nadelhoffer, 2004, 2005; Sousa & Holbrook, 2010), and we refer to this set of findings as the "skill challenge."

In a separate paper, we have analyzed the validity of the side-effect challenge (Guglielmo & Malle, in press), particularly its conditions of occurrence and the components of intentionality it reveals. We concluded that, once comparisons between morally valenced and nonvalenced cases are made truly parallel, and once people can express their judgments in their own terms, the side-effect finding disappears. People are in fact keenly sensitive to the basic components of intentionality that the standard model postulates—belief, desire, and intention—and do not seem to be biased by the moral valence of a side-effect. Moreover, when freely characterizing known but unintended side effects, people do not actually label them intentional, and so they do not make judgments of intentionality without a prior judgment of intention.

The skill challenge, however, still looms. To examine this challenge we will try to identify the conditions under which such findings occur and clarify their implications for theories of intentionality, theories of blame, and for the prospect of valid *mens rea* judgments in the law. For if the negative valence of a defendant's action biases jurors toward "seeing" intent in the action, we would seriously question their fair capacity to assess such intent (Nadelhoffer, 2006b).

1.1. Knobe's original study

Knobe's (2003b) original skill challenge derived from a study in which four components of intentionality (see Fig. 1) were held constant but skill was varied. Together with a manipulation of valence, Knobe employed a 2 (skill: high/low) × 2 (valence: neutral/negative) design. The specific vignettes were as follows: