

# Why Side-Effect Outcomes Do Not Affect Intuitions About Intentional Actions: Properly Shifting the Focus From Intentional Outcomes Back to Intentional Actions

Sean M. Laurent and Brian A. M. Clark  
University of Oregon

Kimberly A. Schweitzer  
University of Wyoming

Over the last decade, many articles have suggested that the “badness” of side-effect outcomes influences perceivers’ intuitions about intentionality, contradicting the traditional notion that mental state inferences lead to moral judgments rather than the reverse. Challenging this assertion, we argue that typically, consideration of intentionality involves thinking about “intentional actions” (things people *do*) rather than unintended outcomes. Across several studies, we offer an explanatory framework describing why side-effect asymmetries emerge. We first establish that people differentiate actions, outcomes, goals, and side effects, associating intentions with goals but intentionality with actions in furtherance of goals, and that each of these components is readily identified in side-effect scenarios. We then demonstrate that when relationships among actions, goals, and side effects are available for consideration in response options, side-effect effects disappear. We additionally show that, because actions are not explicitly referenced, people reinterpret questions about the intentionality of side effects—particularly for harmful outcomes—as asking about intentional *actions* that *caused* side effects, creating a mismatch between participants’ pragmatic and researchers’ literal interpretations. Finally, we demonstrate how harmful side effects shift perceivers’ attention toward considering agents’ knowledge/awareness, whereas beneficial side effects focus attention on intentions/motives, which serves a useful social purpose. We discuss how perceptions of intentionality are not influenced by side-effect valence, although, because of structural differences in how people view harm versus benefit, outcomes influence *which* mental states perceivers consider important when answering questions that are typically asked in side-effects research. Beyond intentionality, we consider how these findings may shed light on trait attribution processes, more generally.

**Keywords:** intentions and intentionality, intentional action, side-effect effects, Knobe effect, moral judgments

**Supplemental materials:** <http://dx.doi.org/10.1037/pspa0000011.supp>

Consider a brief thought experiment. Imagine that John wants to clear the overgrown field behind his house, which abuts his neighbor’s well-loved rose garden. John decides to use herbicide, knowing the poison may also harm this garden. But John does not care about his neighbor’s roses; he just wants to clear the field. So he sprays the poison, and sure enough, the roses are harmed. In this story, although John certainly *knew* that the roses might be harmed, it is not obvious that he *intentionally* harmed them. Instead, the harming was a side effect (SE), resulting from an action (spraying herbicide) in service of a goal (clearing his field). Despite the seeming ambiguity of intentionality here, research has

consistently shown that compared with structurally similar scenarios, in which an agent’s actions bring about a *beneficial* SE—for example, if John’s action helped the roses—people are more likely to say that the agent intentionally caused the harmful outcome (e.g., Knobe, 2003a, 2010b).

This divergence in responses concerning intentionality suggests serious real-world consequences for how people apportion blame. That is, although it makes sense to hold actors responsible for actions that lead to known harmful consequences, calling an outcome intentional simply because it is harmful seems to represent a dangerous flaw in reasoning, because intentionally caused harms are often viewed as more serious than unintentionally caused harms. For example, when a person is killed, ascribing intentionality can mean the difference between a death sentence and life in prison for a defendant found guilty of having caused that death.

A widely used SE scenario describes a Chairman of the Board who is told that starting a new program to increase his company’s profits will result in the helping or harming of the environment, a result about which the chairman does not care (e.g., Knobe 2003a; Knobe, 2004b; Knobe, 2010b). The chairman’s concern is solely to increase profits, so he decides to implement the program, and the

---

Sean M. Laurent and Brian A. M. Clark, Department of Psychology, University of Oregon; Kimberly A. Schweitzer, Department of Psychology, University of Wyoming.

Correspondence concerning this article should be addressed to Sean M. Laurent, Department of Psychology, University of Oregon, 1227 University of Oregon, Eugene, OR 97403. E-mail: [slaurent@uoregon.edu](mailto:slaurent@uoregon.edu) or [seanmlaurent@gmail.com](mailto:seanmlaurent@gmail.com)

environment is subsequently helped or harmed. Participants are then asked whether the chairman intentionally helped or harmed (HH)<sup>1</sup> the environment. Given the helping version, few (e.g., 23%) people say he intentionally helped, but when given the harming version, majorities agree he intentionally harmed (82%; Knobe, 2003b). Similarly, when asked to rate how right/correct it sounds to say the chairman intentionally HH the environment, responses tend to be lower on the scale when the outcome is beneficial versus harmful (e.g., Knobe, 2004a), although in the harming case, ratings are typically only near the midpoint of the scale (see, e.g., Knobe, 2010b; Pettit & Knobe, 2009; see also Wiland, 2007).

Since this intriguing effect was first introduced about a decade ago, a number of articles have generally supported the idea that the morality of a behavior or outcome influences people's intuitions about, among other things, the intentionality of the behavior (e.g., Cushman, Knobe, & Sinnott-Armstrong, 2008; Cushman & Mele, 2008; Knobe, 2003b, 2004b, 2007, 2010a, 2010b; Knobe & Burra, 2006; Nadelhoffer, 2006a, 2006b; Pettit & Knobe, 2009).<sup>2</sup> However, these findings contradict the traditional notion that mental state inferences, such as those regarding intentionality, lead to moral judgments rather than the reverse (e.g., Guglielmo, Monroe, & Malle, 2009; Malle & Nelson, 2003). Because of this, challenges to Knobe's moral influence model have been offered on both methodological and conceptual grounds, generating alternative explanations for this effect (e.g., Adams & Steadman, 2004a, 2004b; Guglielmo & Malle, 2010a, 2010b; Guglielmo, Monroe, & Malle, 2009; Malle, 2006; McGuire, 2012; Scaife & Webber, 2013; Uttich & Lombrozo, 2010; Wiland, 2007).

The current research also challenges the moral influence model. We agree that when presented with certain questions or forced-choice response options, the badness of an SE outcome shifts people's responses in ways consistent with the influence of morality on their intuitions. However, we question whether the goodness or badness of SE outcomes actually affects how participants *think about* intentionality. Put differently, although the SE effect clearly exists, people's responses to relatively simple questions may not always guide us faithfully in the direction of their intuitions, particularly intuitions about concepts as complex as intentionality.

Across several studies, we provide evidence that the badness of an SE does not actually influence intuitions about intentionality. We first show that people differentiate actions, outcomes, goals, and SEs, typically associating the word *intentionally* with actions, but the words *intend* or *intention* with outcomes and goals (and to a lesser extent, with actions), and that people reliably identify each of these elements in the chairman scenario. Next, we provide evidence that, in some ways, help and harm are structurally distinct, making direct comparisons difficult. We then demonstrate that when presented with questions or statements regarding intentional HH, participants think they are being asked about something different from what is ostensibly being asked, highlighting a mismatch between researchers' semantic and participants' pragmatic understanding of questions (see, e.g., Levinson, 1983; Schwarz, 1996). Specifically, we show that participants' understanding of questions about the intentionality of SEs diverges from the literal meaning interpreted by researchers (i.e., regarding the intentionality of SE *outcomes*) and places the emphasis back on the intentionality of *actions*, which are not explicitly referenced but are pragmatically understood to be under consideration. A final set of

studies show that harm relative to help scenarios move people away from thinking about intentional actions and toward thinking about *outcomes brought about* by intentional actions and that consideration of these issues involves reasoning about agents' likely mental states, with harmful SE outcomes shifting people from thinking about an agent's *goals and intentions* to focusing them on the agent's *knowledge and awareness*. These final studies raise interesting theoretical questions about the relative weights perceivers assign to agents' motives (e.g., Reeder, 2009b) versus other mental states when forming impressions of a target person whose actions lead to beneficial versus harmful SE outcomes. That is, person perception may become particularly complex when an agent's actions lead not only to a goal-directed outcome but also to a known but not specifically intended outcome about which the agent does not care.

### The Complexity of Intentionality

Malle (2006) described intentionality as "so deeply ingrained in human cognition that we might count [it] alongside space, time, and causality as one of the fundamental categories with which the mind makes sense of the world" (p. 88). Its importance to social cognition is reflected in its complexity, and part of that complexity is derived from distinctions between actions and outcomes (e.g., causes and effects). Although actions lead to outcomes, and are thus closely related to them, we believe the concepts are relatively distinct. We acknowledge that actions can be identified at multiple levels, from very basic (e.g., moving one's fingers) to those fulfilling higher order goals (e.g., writing a manuscript) (e.g., Vallacher & Wegner, 2012), and that most actions have, as their antecedent, other more basic actions. Despite this, we offer that actions are typically viewed as the things people *do* (whether observable, such as "running," or unobservable, such as "thinking;" see Malle & Knobe, 1997a), often in pursuit of goals. Outcomes, however, are typically viewed as things that *happen*, often, but not always, as a result of an agent's actions.

Admittedly, the line is sometimes blurry. Actions (e.g., sneezing) can be viewed as outcomes in a causal chain (e.g., "The tickle in his nose made him sneeze"). Furthermore, in common usage, actions and outcomes are often referenced simultaneously within a clause. For example, "He closed the door" refers to an action (pushing or pulling the door) and the resultant outcome (the door was closed). This type of statement exemplifies one type of "conversational implicature," where a listener assigns meaning that has not been explicitly referenced, based on cooperative rules of conversation (e.g., Grice, 1989b). These exceptions aside, we advance that outcomes are typically seen as following actions and resulting from them. However, goals are future-oriented and can refer to

<sup>1</sup> Throughout the article, we refer to side effects as *SE* and to word pairings such as "help/harm," "helped/harmed," "helping/harming," or "helping versus harming," as *HH*, where appropriate.

<sup>2</sup> Because the concept of intentionality has been most studied and brought into question regarding SE effects (see, e.g., Knobe, 2010b), we focus primarily on this concept. However, other concepts, such as whether the chairman "decided" to HH the environment, HH "in order to," was "in favor" of HH, "caused" HH, or "desired" HH, have also been noted in the literature (see, e.g., Knobe, 2010b; Pettit & Knobe, 2009). Our arguments throughout, with some adaptation depending on the particular concept being considered, should apply to these other concepts.

desired future actions (e.g., “His goal is to go running”), “focused” desired outcomes (e.g., “His goal is to lose 10 pounds”), or more “diffuse” desired outcomes (e.g., “His goal is to get in shape”). Putting aside more complex cases, we think that starting in early development, humans use cues about goals to help judge whether actions are intentional and assume intentional actions are for pursuit of goals (e.g., Baldwin & Baird, 2001).

Many theoretical models of intentional action describe common elements, including an agent’s knowledge, desires, abilities, and intentions, as well as effort expended in trying to bring about an outcome (e.g., Adams, 1986; Heider, 1958; Jones & Davis, 1965; Shaver, 1985). One model of lay reasoning about intentionality (Malle & Knobe, 1997b) incorporates many of these elements, describing five mental state components—belief, desire, awareness, intention, and skill—that must each be present for an action to be seen as having been intentionally performed. Beliefs regard an agent’s belief that a particular action will lead to a particular outcome, and reference both actions and outcomes by linking potential causes to potential effects. Beliefs do not inform what an agent’s goals might be; simply knowing that an action may lead to an outcome does not suggest one intends to act or desires the outcome. Desire in this model solely regards an agent’s desire for an outcome, and although desire can inform inferences regarding an agent’s goals or reason for acting, it does not necessarily imply a goal but simply represents an agent’s mental state of wanting an outcome to occur. Awareness is often described as referring to whether an agent was aware, while acting, of acting. However, awareness of acting is presumably connected to fulfillment of an intention, in hopes of bringing about a desired outcome. Skill returns to a focus on both actions and outcomes, as it references an agent having the requisite skill to act in a way that brings about a desired outcome. Although many actions can be performed, only those performed with enough skill (often a minimal requirement) to bring about a desired outcome will be considered intentional (see Guglielmo & Malle, 2010b; Malle & Knobe, 1997b).

Although these concepts and the connections between them are complex, intention may be even more so. In the Malle and Knobe model, intention specifically regards an agent’s intention to *act*. Malle and Nelson (2003; see also Malle & Knobe, 2001) argued that intentions properly refer to actions rather than goals or outcomes, because outcomes and goals are in the domain of desires. Theoretically, this makes sense, because people desire outcomes, and may form goals to bring about outcomes through action. However, as we empirically show in several studies, laypeople closely associate the words *intend* and *intention* with an agent’s goals or outcomes that result from an agent’s actions, even while *intentionally* is almost exclusively associated with actions. Thus, unless specifically focused on an agent’s actions, laypeople may equate “intention” or “intend” with “intention to bring about an outcome or achieve a goal.”

Use of the word *intentionality* can also be confusing. Many SE articles specifically reference *intentional action*, but the literature regarding SE effects often conflates actions and SE outcomes. For example, the title of an early article, “Intentional Action and Side Effects in Ordinary Language” (Knobe, 2003a) seems to suggest intentional actions as the topic (see also Knobe, 2003a, 2004b, 2006, 2010b, for similar focus on actions or “behaviors”). However, in this article, discussion consistently centers on whether the chairman intentionally *brought about* HH to the environment (see

also, e.g., Pettit & Knobe, 2009, p. 587), not whether he intentionally acted in a way that led to HH. This may seem like a trivial distinction because outcomes are brought about through actions and are linked to them. However, we think the difference is substantive because although outcomes can result from intentional actions, they can also be brought about indirectly through a causal chain (e.g., an agent intentionally *As* to bring about B, leading to outcome C).

As we empirically show, laypeople think intentionality refers to actions that lead to outcomes and not outcomes brought about by actions. An example might help. It makes less sense to say, “Maria intentionally won the race,” than to say, “Maria intentionally ran as fast as she could (action), in order to win the race (desired outcome).” This is in part because Maria’s actions are controllable; she can decide (i.e., intend) to run as quickly or slowly as she wants. However, winning is not fully under her control, because factors other than her action are involved (e.g., the performance of other runners). Most outcomes share this fate; they can be wished for but never guaranteed. Of interest, actions (running fast) and outcomes (winning) can serve equally well as goals, and Maria can intend to run fast or can intend to win. Despite this, SE outcomes are often called *behaviors* or *actions* (e.g., “side-effect actions”; see Cushman & Mele, 2008; McGuire, 2012), which is arguably imprecise.

We illustrate the idea of intentional outcomes with another example. Malle and Knobe (1997b; Study 1) asked participants to rate the intentionality of several behaviors on an 8-point scale (0 = *not at all*, 7 = *completely*). Some of these behaviors were rated as mostly unintentional (e.g., “Anne is sweating”;  $M = 1.37$ ; p. 104), and others were rated as fairly intentional (e.g., “Anne watered her new plants”;  $M = 6.53$ ). Both of these are things Anne “does.” However, one behavior in this study does not appear to be an action, but an outcome (“Anne got admitted to Princeton”) that presumably resulted from an earlier action (e.g., “Anne applied to Princeton,” “Anne studied hard for the SAT exam”). The average rating for this item ( $M = 3.78$ ) was very close to the midpoint between the lowest (“sweating”) and highest (“watered her plants”) rated items. Although Anne’s admission to Princeton did not seem intentional, it also did not seem unintentional, probably because participants assumed admission was her goal, and that she intentionally acted to increase the likelihood of fulfilling the goal. A similar result (i.e., ratings near the midpoint of an intentionality scale) often emerges when participants are given the harm version of the chairman vignette (see Pettit & Knobe, 2009; Wiland, 2007). This suggests the possibility of two usages of the word *intentional*: a strong version that directly references actions, and a weaker version that indirectly implicates the role of goal-directed actions in bringing about outcomes. If accurate, this still does not suggest multiple intuitions about intentionality. In fact, we think that folk intuitions about intentionality almost exclusively regard actions. However, in some cases, when people say that an outcome was *brought about* intentionally, they may mean that a person *acted* intentionally, which *led to* an outcome, especially when the outcome was foreseen.

### Deconstructing an Ostensibly Simple Scenario

The popular “chairman” scenario (and others like it) may be more complex than is immediately evident. Because of this, trying

to judge the “intentionality” of the SE outcome may be difficult, because it involves complex reasoning about the chairman’s motives and mental states, and trying to disentangle his actions from the outcomes his actions caused. The two versions of the chairman scenario read:

The vice-president of a company went to the chairman of the board and said, “We are thinking of starting a new program. It will help us increase profits, but it will also HH the environment.” The chairman of the board answered, “I don’t care at all about HH the environment. I just want to make as much profit as I can. Let’s start the new program.” They started the new program. Sure enough, the environment was HH.

This simple scenario highlights multiple intentionality components, references an action and two outcomes, and implies a goal. First, beliefs are directly referenced twice in connection with an action, a goal, and outcomes. The chairman knows that starting the program (action) will increase profits (outcome/goal) but that it will also HH the environment (SE outcome). Thus, the chairman enacts the program, knowing the action will have further consequences beyond the desired outcome (see Guglielmo & Malle, 2010a). The chairman’s desire is clear, which makes his goal immediately obvious—increasing profits is the reason for which he acts (see Alvarez, 2009; Knobe, 2007; Wiland, 2007). However, because of how people think about norms, desire is implied in another way (see Uttich & Lombrozo, 2010). Violating norms (e.g., not caring if one’s actions will result in harm) implies more desire than conforming to norms (Guglielmo & Malle, 2010a).<sup>3</sup> Finally, the chairman formed an intention to start the program (or intended to increase profits) and was presumably aware of starting the program, with this action requiring only minimal skill to bring about the intended outcome.

Presented with all of this information, participants probably quickly decide that the chairman did, in fact, act intentionally—he started a program. But, this is not what questions typically ask. Instead, they often ask whether the chairman intentionally helped or harmed the environment. This question about an SE outcome sounds like a question about intentional action, but on the basis of the story, few would assume the chairman went to the woods himself to perform harmful/beneficial actions such as chopping down/planting trees or dumping/cleaning up toxic waste. Similarly, Knobe (2004b; see also Knobe 2006, 2007) asked participants to rate how right or wrong it sounds to say the chairman HH *in order to* increase profits, which attempts to get at whether his action is explained by this reason, because intentional actions should be susceptible to reason explanations. Without digressing to discuss this thorny issue,<sup>4</sup> we point out that this question even more emphatically treats HH as an action. Wiland (2007) argued that which side of the clause “in order to” the action appears is important. When HH was moved from the left to the right side of this clause (i.e., “The chairman started the program in order to HH the environment”), people’s ratings indicated that it sounded quite wrong. However, Wiland’s and Knobe’s questions differ in another way. Wiland’s version properly identifies the chairman’s action (“starting a program”), but ignores his goal, whereas in Knobe’s version, the goal (“increasing profits”) is properly identified, but no action is referenced. Thus, in Wiland’s version, participants probably see the question as “*Was HH the chairman’s goal?*” and in Knobe’s version, they probably see the question as

“*Did the chairman do something, in order to increase profits, which resulted in HH?*”

This is the crux of our argument. When presented with questions regarding intentionality, people expect to be judging the intentionality of actions. However, although results are discussed in terms of actions, the questions typically asked are about SE outcomes, and participants are left with the task of asking themselves: What does this question *mean*? We argue that when questions turn outcomes into seeming actions, participants fill in the blanks themselves, providing their own pragmatic meaning to the questions they are asked. When all actions, outcomes, and goals are properly identified, SE effects should disappear.

## General Methods: Participants

Because a majority of our participants were Internet community samples (i.e., U.S. residents recruited through Amazon’s Mechanical Turk [MTurk] website),<sup>5</sup> we discuss sample selection and exclusion criteria here and provide basic demographic information. Internet data collection occurred in two main waves (see Table 1 for a summary of studies). The first wave (Studies 1–4) took place in early 2013. The second wave (Studies 6–8) occurred mid-summer, 2013. Participation in each wave was tracked via unique MTurk identifiers.

Participants were recruited for three different “sets” of studies in Wave 1. Each set comprised a single brief study or series of studies (some using multiple samples) organized around related questions. Within a set, participants could only participate once (e.g., participation in more than one sample that asked for definitions of concepts such as actions was not allowed). Although participation in more than one set was allowed, few participants did so. To discourage multiple responding, each set used a single IP address, and survey software made multiple responding within sets difficult. When, rarely, more than one response from a single participant within a set was found, data were retained only for the first response completed. Similarly, in Wave 2 (three studies), in the few cases in which participants completed more than one study or one condition in one study, responses for the first study/condition completed were kept, and subsequent responses discarded. Participation in both waves, which were approximately 6 months apart, was allowed. Responses were discarded if data were missing on any variable. Data were also discarded if participants failed to correctly answer a simple attention check question that was included toward the end of each study. Participants were paid between \$0.10 and \$0.35 per study, each of which typically took 3 min or less to complete.

After deleting cases because of multiple responding, missing data, or incorrect responses to attention checks (29 cases), Wave 1

<sup>3</sup> We note that the special case of not *caring* about an outcome indicates ambivalence or lack of concern (i.e., the absence of desire for any specific outcome), perhaps best represented by the midpoint on a scale running from not wanting to wanting an outcome to occur.

<sup>4</sup> We briefly note that we think intentional actions are almost always performed for a reason—to satisfy the agent’s goals. However, sometimes goals may be quite simple or difficult to articulate, such as when the action itself appears to be the desired outcome (e.g., “I did it because I wanted to do it”).

<sup>5</sup> An additional study using undergraduate student participants ( $M_{\text{age}} = 19.74$ ,  $SD = 1.89$ ; 66% female) was collected in spring of 2013 and is reported in Study 5.

Table 1  
*Samples by Wave, Set, and Study/Sample*

Wave	Set	Study and sample	Topic	<i>n</i>
1	1	1.1	Defining actions, outcomes, goals, and side effects; relation of these definitions to intentions and intentionality	98
1	1	1.2	Fill-in-the-blank (words/phrases) for intentions, intentionality, and “in order to” and defining content as actions, outcomes, goals, or side effects	43
1	1	1.3	Leaving intentionality aside, identifying action, goal, and side effect in chairman scenario	79
1	2	2	Rating matched sets of questions regarding benefit and harm	58
1	3	3.1 and 3.2	Identifying intentions and intentional actions in chairman scenario when response options identify action, goal, and side effect (forced choice)	79
1	3	3.3 and 3.4	Identifying intentions and intentional actions in chairman scenario when response options identify action, goal, and side effect (response scale)	79
1	3	4.1–4.6	Meaning of questions regarding “intentional helping/harming”	256
—	—	5	Replication of earlier studies using new (bear-hunter) scenario	65
2	—	6	Abstract helping/harming scenario; intentional vs. unintentional helping/harming; focus on actions vs. goals; reasons and causes	96
2	—	7	Chairman scenario; focus of intentionality on actions vs. outcomes, intentions/goals vs. knowledge/awareness	95
2	—	8	Chairman scenario; focus of intentionality on intentions vs. knowledge, prediction of future behavior	60

*Note.* All samples, except for Study 5, which consisted of undergraduate students, were Internet community samples collected on Amazon’s Mechanical Turk website. Completing more than one study, or participating in more than one sample within a set, was not allowed. Participants could participate in up to three studies in separate sets in Wave 1 (74 participants completed two studies; two participants completed three studies). Participants were also allowed to participate in Wave 1 and Wave 2 (29 participants completed a study in both waves). Wave 1 data collection took place in early 2013. Wave 2 data collection took place in mid-summer, 2013. Study 5 data were collected in spring of 2013. Dashes indicate studies or samples that were not part of a distinct wave or set.

generated 692 cases from 623 unique participants (74 participants completed studies in two sets; two participants completed studies in all three sets). After deleting 48 cases for the same reasons, Wave 2 generated another 251 cases from 222 unique participants (29 participants completed one study in each wave). Across both waves, 845 unique Internet participants took part, generating 943 cases. The full sample, representing participation from people in all 50 states of the United States, had more men (61%,  $n = 515$ ) than women ( $n = 328$ ; two participants reported “other” for gender). Mean age was 29.91 ( $SD = 10.07$ ; range = 18–75). A single question asked, “On most matters (e.g., political, economic, social), where on the following spectrum do you generally consider yourself?” (1 = *Extremely liberal*, 10 = *Extremely conservative*). Responses to this question indicated that the sample leaned liberal, but had substantial variation ( $M = 4.22$ ,  $SD = 2.16$ ).

### Study 1

To make the argument that, through their phrasing, questions concerning the intentionality of SEs unintentionally turn outcomes into seeming actions, we first need to definitively establish several facts—even if these facts may appear self-evident—regarding (a) how people define actions, outcomes, goals, and SEs; (b) how these terms are typically related to the concepts of intentions and intentionality; and (c) how people easily recognize each of these elements in a widely used SE scenario. This is important because many of our later claims hinge on people’s understanding of these terms, the relation of these terms to intentions and intentionality, and people’s ability to recognize these concepts when responding to questions about the intentionality of SEs. Thus, to strengthen later claims, it seems prudent to lay a sound, if simple, empirical foundation before moving on to questions of greater complexity.

In Study 1, using three different samples, we examine several related questions regarding (a) whether people see actions, goals, outcomes, and SE outcomes as distinct and clearly defined; (b) how most people think the words *intend* and *intention* refer to agents’ goals, but that *intentionality* refers to actions rather than outcomes or goals (and not to SE outcomes); and (c) whether, leaving questions of intentionality aside, participants presented with the chairman story agree in their identification of his action, the goal of his action, and the SE that resulted from his action (e.g., Gintis, 2010).

### Method

**Sample 1.1.** In two conditions, participants ( $n = 98$ ) were asked to define actions, outcomes, goals, and SEs, and to apply these definitions to their understanding of intentions and intentionality.<sup>6</sup> In both conditions, participants were asked (in separate questions): “An (action, outcome, goal, SE) is . . .” Response options were (a) something someone *does*; (b) a *result*, sometimes intended and sometimes not, brought about by something someone did; (c) something a person wants to *accomplish* or *an end* toward which effort is directed; and (d) an unplanned *secondary* result, often brought about by someone’s action, that is not usually intended, but is sometimes foreseen. In an “abstract” condition, this question was followed with two separate prompts regarding the meaning of intentions and intentionality: “If you heard that (someone intentionally X’d; A person intended to X), X probably refers to . . .” In a “concrete” condition, participants responded to the prompts “Annie intentionally \_\_\_\_\_,” and “Annie intended to

<sup>6</sup> In all studies reported in this article, conditions were between subjects, with random assignment to condition.

\_\_\_\_\_.” Response options in both conditions were essentially the same as above (i.e., referring to actions/things people do, outcomes/things that happen as a result of actions, goals/things people try to accomplish, and SEs/unplanned secondary results), phrased accordingly.<sup>7</sup>

**Sample 1.2.** Participants ( $n = 43$ ) typed in words or phrases to complete simple sentences. The agent varied by condition (“John Smith” [JS]; “The chairman of the board” [CB]). Three sentences were completed. These were (a) “JS/CB intended to \_\_\_\_\_” (b) “JS/CB intentionally \_\_\_\_\_” and (c) “JS/CB \_\_\_\_\_ in order to \_\_\_\_\_.” Following this, participants selected response options that best captured the essence of their self-created content. Options for each blank were (a) an action performed by JS/CB or something he did, (b) an outcome deliberately brought about by JS’s/CB’s action or that resulted from his action, (c) a goal that JS/CB wanted to fulfill or something he wanted to happen, or (d) an unplanned outcome that resulted from JS’s/CB’s action. Tenses and phrasing of response options varied grammatically as a function of question (e.g., *did* vs. *planned to do*).

**Sample 1.3.** Participants ( $n = 79$ ) read either the helping or the harming version of the chairman vignette and were asked, in separate questions: “In the story above, (the chairman’s action; goal; a SE of the chairman’s action) was . . .” Response options for all questions were (a) starting a new program, (b) increasing profits, and (c) HH the environment. In each condition, half the participants were provided definitions of actions, goals, and SEs, and the other half were not.

## Results

**Samples 1.1 and 1.2.** Chi-square independence tests were used to examine whether responses to questions (about actions, outcomes, goals, and SEs in Sample 1.1; about intentions and intentionality in Samples 1.1 and 1.2; about doing \_\_\_\_\_ in order to \_\_\_\_\_ in Sample 1.2) differed as a function of conditions (i.e., abstract vs. concrete in Sample 1.1; JS vs. CB in Sample 1.2), and whether Samples 1.1 and 1.2 differed in their response to questions regarding intentions and intentionality. No significant condition-based differences emerged in either sample in response to any question, and no significant sample-based differences were found in response to questions regarding intentions and intentionality.

Table 2 provides frequencies and percentages of responses to each question in Samples 1.1 and 1.2. In Sample 1.1, most people defined actions as “things people do” (92.9%), outcomes as “results brought about by things people do” (91.8%), goals as “ends toward which effort is directed” (96.9%), and SEs as “unplanned secondary effects” (91.8%). Across both samples, a majority of participants thought the word *intended* referred to an agent’s goals (60%), with the remainder dividing their definitions between actions (20.5%), outcomes (16%), and SEs (3.5%). However, across both samples, a majority of participants thought the word *intentionally* referred to actions or something an agent did (71%), with the remainder focusing definitions on outcomes (15.5%), goals (10.5%), and SEs (3%). In Sample 1.2, participants also filled in blanks in response to the following prompt: “(JS/CB) \_\_\_\_\_ in order to \_\_\_\_\_.” By their own reckoning and using their own examples, almost all participants (88.3%) thought that agents *did* something (i.e., performed an action; e.g., “voted” or “encouraged his employees”) in order to *bring about an outcome* (46.5%; e.g.,

“save his job”) or *achieve a goal* (41.9%; e.g., “hit their sales targets”).

**Sample 1.3.** HH had no significant effect on any variable. However, participants who were not given definitions (9/37; 24.3%) were more likely than those given definitions (0/42; 0%) to call “increasing profits” the chairman’s action,  $\chi^2(1, N = 79) = 11.53, p = .001$ . All other participants indicated his action was starting a program, including all participants in the definition condition. The definition condition had no effect on selection of responses related to goals or SEs. Almost all participants (78/79) indicated that increasing profits was the chairman’s goal. One participant in the helping condition thought his goal was helping the environment. All participants (i.e., 79/79) identified the HH of the environment as an SE.

## Discussion

Study 1 provides foundational support for several key arguments. First, it establishes that people typically agree about what actions are (the things people *do*) and see them as distinct from outcomes (things *that result* from actions), goals (things people want to *achieve*), and SEs (*unplanned secondary outcomes* that might or might not be foreseen). Second, it shows that although some people associate the word *intention* with actions or outcomes, most people use the word to refer to goals. Unlike intentions, most people associate *intentionality* with actions, and whereas some participants (approximately 26%) associated intentionality with either outcomes or goals, almost no participants (4/141 across Samples 1.1 and 1.2) associated it with SE outcomes. In addition, Sample 1.2 showed that people think agents *act* in order to *achieve goals* or *bring about outcomes*, but do not bring about SEs to achieve other outcomes or goals.

The findings concerning intentions and intentionality emerged when questions were framed using abstract concepts (e.g., “If you heard that someone intentionally ‘Xd,’ ‘X’ most likely refers to”) or concrete examples (“Annie intentionally \_\_\_\_\_”) and when people used their own words and then defined the concepts their words represented. This replication across multiple samples and question formats suggests that when people hear or see a statement such as “The chairman intentionally,” they expect an action to follow (e.g., “started a program”), and not an SE outcome (e.g., “HH the environment”).

Consistent with this, when people read the chairman scenario (Sample 1.3), most participants identified his action as “starting a program”—although when not given definitions, a minority labeled his action as “increasing profits.” Thus, despite the HH in the chairman story often being referred to as an action (e.g., Cushman & Mele, 2008; Pettit & Knobe, 2009), no participants identified it as such in either version of the vignette; instead, all participants saw the HH as an SE of the action, performed in connection with the goal of increasing profits. This provides evidence that most people correctly (in our view) identify each element of the story when intentionality is not specifically referenced.

<sup>7</sup> A complete description of all studies, including all instructions to participants, can be found in the Appendix, available in the supplemental material.

Table 2  
Frequency and Percentage of Responses in Study 1

Terms/questions	Response options: Frequency (%)			
	Something someone does	A result often brought about by something someone does	An end toward which effort is directed	An unplanned secondary effect, sometimes foreseen
Sample 1.1 ( $n = 98$ ): Define terms/answer questions using response options				
An action is . . .	91 (92.9%)	2 (2.0%)	4 (4.1%)	1 (1.0%)
An outcome is . . .	0 (0.0%)	90 (91.9%)	5 (5.1%)	3 (3.0%)
A goal is . . .	0 (0.0%)	2 (2.0%)	95 (97.0%)	1 (1.0%)
A side effect is . . .	1 (1.0%)	6 (6.1%)	1 (1.0%)	90 (91.9%)
	Do/did something	Bring/brought about outcome deliberately	Achieve goal	Bring/brought about side effect
Intended to _____	22 (22.4%)	19 (19.5%)	55 (56.1%)	2 (2.0%)
Intentionally _____	71 (72.5%)	15 (15.3%)	10 (10.2%)	2 (2.0%)
Sample 1.2 ( $n = 43$ ): Fill-in-the-blank and define content using response options				
Intended to _____	7 (16.3%)	4 (9.3%)	29 (67.4%)	3 (7.0%)
Intentionally _____	29 (67.4%)	7 (16.3%)	5 (11.6%)	2 (4.7%)
Person _____ in order to _____	38 (88.3%)	3 (7.0%)	2 (4.7%)	0 (0.0%)
	4 (9.3%)	20 (46.5%)	18 (41.9%)	1 (2.3%)

*Note.* In Sample 1.1, participants defined the given terms by selecting one of the provided response options. Participants in Sample 1.2 completed sentences (e.g., John Smith intended to \_\_\_\_\_; John Smith intentionally \_\_\_\_\_; John Smith \_\_\_\_\_ in order to \_\_\_\_\_) by typing responses into blank fields. Following this, they defined the content they created using one of the provided response options.

## Study 2

Study 2 addresses a different, but important, issue before we return to an examination of the relations of actions, outcomes, goals, and SEs to intentions and intentionality. Research has suggested that people are blamed more for doing “bad” things than they are praised for doing “good” things and that this blame/praise asymmetry may serve a useful social function (e.g., Malle & Nelson, 2003). Morally good behavior may seem uninformative because it reflects normative, socially prescribed behavior. However, bad, norm-violating behavior, because it seems predictive of future behavior, is socially useful (e.g., Bartsch & Young, 2010; Uttich & Lombrozo, 2010; Ybarra, 2002).

These differences in how people view good and bad behavior may also be reflected in structural differences in language use when discussing good and bad outcomes that are not pursued as goals (e.g., SEs). As a class, unplanned but foreseen secondary outcomes are passively “allowed” rather than actively “performed” (see Cushman, Knobe, & Sinnott-Armstrong, 2008, for a related discussion). However, although the concept of “allowing” harm sounds reasonable (e.g., for negligent actions; see Nuñez, Laurent, & Gray, 2014), allowing benefit or help sounds strange, because normatively, people should “encourage” benefit. Similarly, people “cause” and are “held responsible for” harmful but not helpful actions. These differences may contribute to the asymmetries found in SE studies.

## Method

Participants ( $n = 58$ ) were asked four matched sets of questions regarding HH (i.e., one question in each set asked about helping and the other about harming), using 7-point response scales ( $-3 =$  Sounds wrong/strange,  $0 =$  In between, and  $+3 =$  Sounds right/normal). Questions 1 and 2: “When a person’s actions HH some-

one, the person should be held responsible for this HH.” Questions 3 and 4: “A person allowed HH to come to the environment.” Questions 5 and 6: “A person allowed the environment to be HH.” Questions 7 and 8: “A person caused HH to the environment.” Three additional sets of questions also asked for participants’ agreement with three matched statements on a 7-point scale ( $-3 =$  Totally disagree,  $0 =$  Neither disagree nor agree,  $+3 =$  Totally agree). Questions 9 and 10: “People have a responsibility to encourage benefit to come (prevent harm from coming) to the environment.” Questions 11 and 12: “When a person does not act to encourage help (prevent harm) (e.g., to the environment), the person is negligent.” Questions 13 and 14: “A person should be considered ‘responsible’ if their company starts a program that unintentionally but considerably helps (negligently harms) the environment.”

## Results and Discussion

Paired  $t$  tests were used to examine response differences in all question pairs, followed by one-sample  $t$  tests testing responses against the midpoint of the scale for “wrong/right” questions (1–8) only. Significant differences were found in responses to the HH versions of all questions,  $t(57)$  from 2.15 to 5.98,  $ps$  from .04 to  $< .001$ ,  $ds$  from .40 to 1.16; the question pair asking about a person’s responsibility if her or his company starts a program that considerably helps (negligently harms) the environment was marginally significant,  $t(57) = 1.91$ ,  $p = .06$ ,  $d = .25$  (see Table 3). Single-sample  $t$  tests also showed that “harm” versions of Questions 1–8 sounded right/normal, with  $ts$  ranging from 2.95 to 13.53,  $ps \leq .005$ , whereas help versions either sounded “in between” (Questions 1, 3, and 5,  $ns$ ) or wrong (Question 7),  $t(57) = 3.58$ ,  $p = .001$ .

These findings highlight how structural differences in how people talk about harming and helping may contribute to SE

Table 3  
*Study 2 Means, Standard Deviations, and Associated Statistics*  
 ( $n = 58$ )

Statements	Help $M$ ( $SD$ )	Harm $M$ ( $SD$ )	$t$	$p$	$d$
1 and 2	0.33 (1.85)	1.98 (1.12)	5.98	<.001	1.08
3 and 4	-0.43 (2.19)	1.02 (1.81)	3.45	.001	0.72
5 and 6	-0.14 (1.97)	0.76 (1.96)	2.15	.036	0.46
7 and 8	-1.00 (2.13)	1.43 (2.06)	4.70	<.001	1.16
9 and 10	1.09 (1.30)	1.60 (1.31)	3.09	.003	0.40
11 and 12	0.00 (1.51)	0.93 (1.14)	4.60	<.001	0.70
13 and 14	0.60 (1.34)	0.97 (1.53)	1.91	.062	0.25

*Note.* The response scale for Statements 1–8 was  $-3 =$  *Sounds wrong/strange*,  $0 =$  *In between*,  $+3 =$  *Sounds right/normal*; for Statements 9–14, the response scale was  $-3 =$  *Totally disagree*,  $0 =$  *Neither disagree nor agree*,  $+3 =$  *Totally agree*. Statements 1 [2]: When a person's actions help [harm] someone, the person should be held responsible. Statements 3 [4]: A person allowed help [harm] to come to the environment. Statements 5 [6]: A person allowed the environment to be helped [harmed]. Statements 7 [8]: A person caused help/harm to the environment. Statements 9 [10]: People have a responsibility to encourage benefit to come [prevent harm from coming] to the environment. Statements 11 [12]: When a person does not act to encourage help [prevent harm] (e.g., to the environment), the person is negligent. Statements 13 [14]: A person should be considered "responsible" if their company starts a program that unintentionally but considerably helps [negligently harms] the environment.

effects, such as those concerning causality and "allowing" outcomes to occur, particularly when SEs are framed as actions. People think it much stranger to talk about "causing" help, being "responsible for" help, or "allowing" help than to use the same words in reference to harm (which sounds "right"). Furthermore, participants think people have a greater responsibility to avoid harming the environment than to benefit it. Last, participants saw company owners as somewhat more responsible for harm than for benefit and more negligent when they cause harm than when they fail to encourage benefit. In sum, these differences show that help and harm may not be easily equated, because although helping is somewhat good, harming seems pretty bad. This suggests that passively allowing a beneficial outcome to occur is not notable, whereas allowing a harmful outcome is socially meaningful.

### Study 3

Study 3 returns the focus to actions, goals, and outcomes, and their relation to intentions and intentionality. Having first established in Study 1 (Sample 1.3) that without mentioning intentionality or intentions, majorities identify the chairman's action as starting a program, his goal as increasing profits, and the SE as HH the environment, we wanted to know whether (a) people view the chairman's *intentions* as relating to his action or his goal (or perhaps, to multiple goals); and (b) people would see his action, the SE, or both as his intentional action, performed in service of a goal. More generally, we wanted to see how participants would respond to questions of intention and intentionality when the chairman's action, goal, and the SE were simultaneously available for consideration in response options. We expected that when able to fully evaluate and consider each of these elements, few participants would say that he intentionally HH the environment but that most would say he intentionally started a program to increase profits. Thus, no asymmetries should emerge.

### Method

Four separate samples were used, and all participants read either the help or the harm version of the chairman vignette. After reading the vignette, Samples 3.1 and 3.2, respectively, responded to related sets of forced-choice questions, each with multiple response options. Rather than using forced-choice options, Samples 3.3 and 3.4 rated each response option given, respectively, to Samples 3.1 and 3.2 on a 7-point scale ( $-3 =$  *Sounds wrong/inaccurate*,  $0 =$  *In between*,  $+3 =$  *Sounds right/accurate*).

Participants in Sample 3.1 ( $n = 39$ ) were asked two questions: "Which of the following statements most accurately describes . . . (the chairman's intentions) (what the chairman did intentionally)?" Response options to the first question were "The chairman's intention was . . ." (a) "to start a program," (b) "to increase profits," and (c) "to help (harm) the environment." Response options to the second question were "The chairman intentionally . . ." (a) "started a program, in order to increase profits," (b) "started a program, in order to help (harm) the environment," and (c) "helped (harmed) the environment, in order to increase profits." Participants in Sample 3.3 ( $n = 40$ ) independently rated each of these response options on the scale described above.

Participants in Sample 3.2 ( $n = 40$ ) were asked two similar questions: "Which of the following statements most accurately describes . . . (the chairman's goals, if he had any) (the chairman's intentional action or actions)?" Response options to the first question were "In order to . . ." (a) "increase profit, the chairman started a new program," (b) "HH the environment, the chairman started a new program," and (c) "increase profit and HH the environment, the chairman started a new program." Response options to the second question were "The chairman's intentional action(s) . . ." (a) "was to HH the environment," (b) "was to start a program to increase profits," and (c) "were both to HH the environment and to start a program to increase profits." Participants in Sample 3.4 ( $n = 39$ ) independently rated each of these response options on the scale described above.

### Results

**Samples 3.1 and 3.2.** In both samples, chi-square independence tests were used to examine whether HH condition affected responses to the forced-choice questions. No significant condition-based differences in response preference were found. Separate chi-square tests testing the null hypothesis of equal frequencies across response options revealed significant clustering in participants' selection of response options to all questions in both samples, with  $\chi^2$  (all  $df = 2$ , Sample 3.1  $N = 39$ ; Sample 3.2  $N = 40$ ) statistics ranging from 31.41 to 36.10 (all  $ps < .001$ ). Table 4 summarizes the frequency that participants in both samples selected each response option, organized by HH condition. Almost every participant selected "increasing profit" to describe the chairman's intention (Sample 3.1) or goal (Sample 3.2). This was true even when participants had the chance to identify increasing profits *and* helping/harming as dual goals (Sample 3.2). Similarly, almost all participants identified what the chairman "did intentionally" (Sample 3.1) or his "intentional action" (Sample 3.2) as starting a program to increase profits. This was true even though participants could have chosen HH as his intentional action, HH as what he did to increase profits (Knoke, 2004b), or starting a program as the action performed in order to HH the environment

Table 4  
*Forced-Choice Responses to the Chairman's Intention and What He Did Intentionally (Sample 3.1) and His Goal and Intentional Action (Sample 3.2) in Study 3*

Study 3: Help/harm forced choice response frequencies (%)				
Sample 3.1 ( <i>n</i> = 39)	Intention	Start a program 2 (11%)/0 (0%)	Increase profits 16 (89%)/21 (100%)	Help/harm environment 0 (0/0%)/0 (0%)
	Intentionally did	Program, for profits 18 (100%)/20 (95%)	Program, to help/harm 0 (0%)/0 (0%)	Help/harm, for profits 0 (0%)/1 (5%)
Sample 3.2 ( <i>n</i> = 40)	Goal	To increase profits 19 (95%)/19 (95%)	To help/harm 0 (0%)/0 (0%)	To do both 1 (5%)/1 (5%)
	Intentional action	To start program 20 (100%)/19 (95%)	To help/harm 0 (0%)/0 (0%)	Both 0 (0%)/1 (5%)

*Note.* Participants in both samples read either the help or the harm version of the chairman vignette, and selected one of the response options for two questions (Sample 3.1: intention and intentionally did; Sample 3.2: goal and intentional action).

(Wiland, 2007). Furthermore, although given the opportunity to label both starting the program and HH as intentional actions furthering the goal of increasing profits (Sample 3.2), only one participant selected this option. Thus, using forced-choice responses, no HH asymmetries were found in how participants viewed intentionality.

**Samples 3.3 and 3.4.** Participants' ratings of the same sets of response options used, respectively, for Samples 3.1 and 3.2 in Samples 3.3 and 3.4 showed a similar pattern (see Table 5). Data were analyzed using four separate (two analyses for each sample) 2 (help vs. harm condition)  $\times$  3 (intention/goal or intentionality/intentional action questions) mixed analyses of variance (ANOVAs), with repeated measures on the second factor. Repeated contrasts (i.e., Question 1 vs. Question 2 and Question 2 vs. Question 3) were also examined in each analysis to explore whether there were any mean differences in response to different questions.

In Sample 3.3, main effects of HH condition emerged for both intention and intentionality questions (respectively,  $F_s[1, 38$  in Sample 3.3;  $1, 37$  in Sample 3.4] = 17.34 and 7.91,  $p_s < .01$ , partial  $\eta^2$ s = .31 and .17). For both question sets, these main effects reflected higher average ratings across questions in the harm condition than in the help condition. In Sample 3.4, the main effects of condition were not significant. Critically, there were no significant omnibus interactions between HH and questions in any analysis, which would be predicted by the moral influence model. Instead, in each analysis, there were large main effects of questions, with  $F_s(2, 76$  in Sample 3.3;  $2, 74$  in Sample 3.4) ranging from a low of 58.61 (partial  $\eta^2 = .61$ ) for intention questions in Sample 3.3 to a high of 143.38 (partial  $\eta^2 = .80$ ) for goal questions in Sample 3.4 (all  $p_s < .001$ ), showing substantial variability in responses to different questions. In each analysis, all repeated contrasts were significant ( $p_s \leq .001$ ).<sup>8</sup> Fully consistent with the frequencies of response selection in Samples 3.1 and 3.2, statements about increasing profit were always seen as most accurate for intention and goal questions, and starting a program was always seen as the most accurate description of what the chairman did or his intentional action, in connection with this goal.

These analyses were followed with single-sample  $t$  tests (collapsed across HH) to examine whether responses to each question

in Samples 3.3 and 3.4 were significantly different from the midpoint of the scale (which indicated that a statement sounded neither accurate nor inaccurate). With the exception of one statement in Sample 3.3—that the chairman intentionally HH the environment, in order to increase profits ( $p = .17$ )<sup>9</sup>—responses indicated that statements sounded either wrong or right,  $t_s(39$  in Sample 3.3;  $38$  in Sample 3.4) ranged from 2.69 to 21.94,  $p_s \leq .01$ . Overall, there was strong agreement that the chairman's intention/goal was to increase profits, somewhat weaker agreement that his intention/goal was to start a program, and strong disagreement that his intention or goal was to HH the environment. Similarly, in connection with this intention/goal, there was strong agreement that his intentional action and what he did intentionally was to start a program, and no agreement that any other statement was accurate (i.e., qualified as intentional).

A final set of tests examined responses to all 12 questions as a function of condition alone (i.e., HH). Significant differences emerged on two questions only, in Sample 3.3. The first was whether it was the chairman's intention to HH the environment; participants given the harm version thought this statement sounded more accurate than those given the help version,  $t(38) = 3.69$ ,  $p < .001$ ,  $d = 1.19$ . Similarly, an effect emerged for the statement suggesting that the chairman intentionally HH the environment, in order to increase profits,  $t(38) = 2.14$ ,  $p = .04$ ,  $d = .68$ . Again, this statement sounded more inaccurate in the help condition.

<sup>8</sup> In one analysis, outcome interacted ( $p = .04$ ) with a contrast testing response differences between "The chairman's intention was to increase profits" and "The chairman's intention was to HH the environment" (Sample 3.3). In both conditions, differences between ratings for these questions was significant ( $p_s < .001$ ), with intention to increase profits seen as more accurate than intention to HH. However, the mean difference between ratings for these two questions was larger in the help condition ( $M = 4.84$ ) than in the harm condition ( $M = 3.29$ ).

<sup>9</sup> Further analyses showed that for this question, in the help condition, average response was significantly below the scale midpoint,  $t(18) = 2.69$ ,  $p = .02$ . In the harm condition, average response was not significantly different from zero.

Table 5  
*Means and Standard Deviations for Ratings of the Same Forced-Choice Response Options Given to Samples 3.1 and 3.2 by, respectively, Samples 3.3 and 3.4, Organized by Helping/Harming Conditions (Study 3)*

Variable	Help	Harm
	<i>M (SD)</i>	<i>M (SD)</i>
Response options given to Sample 3.1, rated by Sample 3.3 ( <i>n</i> = 40)		
The chairman's intention was . . .		
to start a program	0.95 (2.37)	1.95 (1.83)
to increase profits	2.42 (1.30)	2.90 (0.30)
to help/harm the environment	-2.42 (1.17)	-0.38 (2.13)
The chairman intentionally . . .		
started a program, in order to increase profits	2.26 (1.73)	2.81 (0.51)
started a program, in order to help/harm the environment	-2.42 (1.07)	-1.86 (1.71)
helped/harmed the environment, in order to increase profits	-1.26 (2.05)	0.19 (2.23)
Response options given to Sample 3.2, rated by Sample 3.4 ( <i>n</i> = 39)		
In order to . . . , the chairman started a new program		
increase profits	2.78 (0.73)	2.62 (0.80)
help/harm the environment	-2.38 (1.54)	-1.95 (1.60)
increase profits and help/harm the environment	-0.67 (1.50)	-0.95 (2.22)
The chairman's intentional action(s) was (were both) to . . .		
help/harm the environment	-2.50 (1.29)	-1.76 (1.70)
start a program to increase profits	2.67 (1.19)	2.81 (0.51)
help/harm and to start a program to increase profits	-1.67 (1.64)	-0.90 (1.97)

*Note.* Participants in both samples read either the help or harm version of the chairman vignette. Participants in Sample 3.3 and 3.2 respectively rated all response options given to participants in Samples 3.1 and 3.2. All questions were rated on 7-point scales (-3 = *Sounds wrong/inaccurate*, 0 = *In between*, +3 = *Sounds right/accurate*).

## Discussion

Study 3 provides compelling evidence that the badness of SE outcomes does not actually influence lay intuitions regarding intentions and intentionality, at least as applied to a common SE scenario. They also suggest that responses to questions of intentionality index something more fundamental than which side of "in order to" the HH occurs (Wiland, 2007). That is, although we agree that phrasing of questions is important, what is most important is providing participants with response options that properly identify goals, actions in pursuit of goals, and SE outcomes.

These results show that participants thought the chairman did only one thing intentionally: He started a program, to pursue his goal of increasing profits. Specifically, in Samples 3.1 and 3.2, participants overwhelmingly selected "increasing profits" as the chairman's intention/goal, and "starting a program" as his intentional action, completed in pursuit of this goal. This was true even though they could have agreed that HH was his intentional action, that he HH to increase profits, that he started the program to HH, or that he both HH and started the program to increase profits. On rating scales (Samples 3.3 and 3.4), a similar pattern emerged. Participants thought it sounded very accurate to say he intentionally started a program to increase profits, whereas it sounded quite inaccurate to say he intentionally HH the environment, or HH to increase profits. These findings are fully consistent with the findings in Study 1, which showed that although some people associate intentionality with outcomes brought about through actions (e.g., to achieve goals), most people associate intentionality with the actions themselves, and almost no people think intentionality refers to bringing about secondary outcomes (i.e., SEs).

Although the HH of the environment is an outcome, it is a *secondary* outcome and was not the chairman's goal. And because it is not an action, it does not really qualify as something the chairman intentionally "did."

For two questions (but not the remaining 10), responses were influenced by HH outcome. We think it is important to note that although outcome affected responses on these questions, means were not significantly different from zero in the harm conditions. Although arguments have been advanced suggesting that the actual values for ratings of intentionality are not particularly important (e.g., such as for a scale going from "sounds wrong" to "sounds right") and that the important finding is the differences in ratings for harm versus help (e.g., Knobe, 2010b; Pettit & Knobe, 2009), we do not agree. If we are to trust this rating scale, then ratings in the middle mean that a statement sounds neither right/accurate nor wrong/inaccurate. As a point of comparison, when asked about the intentionality of starting a program, average ratings were near the ceiling of agreement, showing that it sounds quite right to say this was what the chairman did intentionally. However, it would be unwise to dismiss the difference in ratings that are so often found (and found for two questions here) as meaningless. Our remaining studies attempt to address why, aside from structural differences in how people evaluate HH outcomes, robust SE findings emerge at all, particularly if intentionality is about actions and SEs are simply unintended but foreseen and allowed outcomes.

## Study 4

Questions matter. Participants' responses to simple and seemingly straightforward questions about a complicated topic may not

reveal the full intricacy of their thinking about the concept. Instead, their responses may represent a best effort to answer the question they think is being asked by assuming a question means something that has not been explicitly stated. This is not a new idea: As discussed by Schwarz (1996) and others, the principles of conversational logic and norms (e.g., Grice, 1989a, 1989b) can lead participants to infer meaning that is not explicitly present from a researcher's questions or statements, based on assumed common ground (e.g., Levinson, 1983). That is, when the literal meaning of a researcher's question or statement is somewhat ambiguous or unclear, participants will try to infer the meaning pragmatically. For example, imagine that a research participant reads about Sam, who, to console a friend, buys the friend a present. If asked, "Did Sam intentionally comfort his friend?" the participant would probably, even if unaware of doing so, answer yes and correctly reinterpret this as meaning, "Did Sam intentionally buy his friend a present, in order to comfort him?" In this example, the reinterpretation is not problematic because this is likely what the researcher meant by the question. Furthermore, because the action (buying a present) was in furtherance of the desired goal/outcome (comforting a friend), intent and intentionality are fully aligned.

However, when responses to a question, as asked, are interpreted literally by a researcher when the participant has understood the question to mean something else, problems might arise. Similarly, when available response options to a question do not capture the essence of how participants think about an issue, they may select a "best" response with which they do not fully agree, or reinterpret the meaning of the available options to align with their intuitions. Guglielmo and Malle (2010a) made this point by asking participants to choose between whether the chairman intentionally (a concept that most would agree is complex and multifaceted) or knowingly (a simpler concept that itself contributes to intentionality judgments) harmed the environment. Most participants chose the latter. However, "knowingly harming" still turns a passively allowed SE outcome (harming) into an intentionally committed action, suggesting harming was the thing the chairman knowingly did, rather than brought about through an action.

We argue that when asked questions such as "Did the chairman intentionally HH the environment?" participants' answers may be to a different, *unasked* question. For example, if participants think this question asks whether the chairman intentionally *did something* that resulted in HH, and if "starting a program" is what they think he did intentionally, responses to this question take on a new meaning. That is, if people assume the question is asking about an action rather than the stated SE outcome, and people agree what the intentional action is, it suggests—in a way that complements the findings of Study 3—that judgments of intentionality are not influenced by outcome. Additionally, if more fully explicated response options result in HH asymmetries (depending on the asymmetries), this leads to the chance for further explanation of and insight into why SE effects emerge. However, even if no asymmetries are found, if questions are understood to mean something different than what is explicitly stated, we would still learn that the answers typically received may not have been to the questions asked, representing an important limitation of prior work. Furthermore, this would increase our confidence that SE asymmetries do not reflect differences in how people judge intentionality itself, but instead are

based on other core differences in how people evaluate harmful versus helpful SE outcomes.

## Method

Six samples, after reading either the help or the harm version of the chairman vignette, were asked about the meaning of statements and questions like "The chairman intentionally HH the environment" or "Did the chairman intentionally HH the environment?" using forced-choice response options. These questions and response options are described below.

**Sample 4.1.** After reading the vignette, participants ( $n = 41$ ) were asked to consider the statement "The chairman intentionally HH the environment" and to answer two questions, each requiring a yes or no response: "If you think this statement is implying that the chairman's intentional action was to HH the environment, answer yes. If you think the statement actually means something different, answer no" and "Do you think this statement actually means 'The chairman intentionally started a program he knew would result in benefit (harm) to the environment.'"<sup>10</sup>

**Sample 4.2.** After considering the same statement given to Sample 4.1, participants ( $n = 43$ ) were asked to choose from two response options: "This statement suggests the . . ." a) "intentional action was to HH the environment" and b) "HH was a SE of a different intentional action."

**Sample 4.3.** Following the vignette, participants ( $n = 45$ ) were asked to consider the question "Did the chairman intentionally HH the environment?" They then selected from the following options: "This question is asking if the . . ." a) "chairman's intentional action was to HH the environment," or b) "HH was a known SE of an intentional action."

**Sample 4.4.** After reading the vignette, participants ( $n = 42$ ) considered the same question given to Sample 4.3 and were then asked to choose from one of two response options: "This question is asking if the chairman's intentional action was to . . ." a) "HH the environment," or b) "start a program that he knew would HH the environment."

**Sample 4.5.** After reading the vignette, participants ( $n = 43$ ) considered the same statement given to Samples 4.1 and 4.2. They were then asked to select from one of several response options to indicate their understanding of the meaning or gist of the statement. Response options were "The statement means the chairman intentionally . . ." a) "HH the environment," b) "started a program to increase profits, knowing it would HH the environment," or c) "increased profits, knowing it would HH the environment."

**Sample 4.6.** After reading the vignette, participants ( $n = 42$ ) considered the statement "In order to increase profits, the chairman intentionally HH the environment" and were then asked, "What does this statement really mean?" Response options were (a) "The chairman intentionally HH the environment in order to increase profits" and (b) "In order to increase profits, the chairman intentionally started a program he knew would HH the environment."

<sup>10</sup> The instructions and response options presented in text for Study 4 are slightly modified and abbreviated from the original to conserve space, but fully convey what participants were asked. The full text of all instructions and questions can be found in the Appendix, offered in the supplemental material.

## Results and Discussion

Chi-square independence tests were used to examine whether condition (i.e., helping vs. harming) or question (Sample 4.1 only) was associated with response in each sample. For those samples in which no significant association was found, chi-square tests of the null hypothesis of equal responses across response options (collapsed across condition) were conducted. Table 6 reports the frequencies with which participants selected each possible response in each sample, organized by helping versus harming outcome.

Significant associations between condition and response option were found to the second question in Sample 4.1, and for Samples 4.3 and 4.4,  $\chi^2$ s ( $df = 1$ ,  $N$ s = 41, 45, and 42, respectively) ranged from 5.31 in Sample 4.1 (Question 2) to 9.75 in Sample 4.3 (all  $ps \leq .02$ ). A marginal association between condition and response option was also found in Sample 4.2,  $\chi^2(1, N = 43) = 2.79, p = .095$ . In each case, a greater number of participants presented with the harming version thought the questions/statements meant something *other than* what was written, compared with those who read the helping version. For example, in Sample 4.3, 83.3% (20/24) of participants given the helping version thought the question “Did the chairman intentionally help the environment?” was asking whether the chairman’s *intentional action* was to help the environment. Only 38.1% (8/21) of participants in the harming version thought so. The remainder responded that the question was asking

whether the harming was a known SE of an intentional action. These asymmetries demonstrate how majorities in both helping and harming groups (i.e., in past research) might give “correct” but different answers. That is, the chairman’s intentional action was not to help the environment, but to start a program, and the harming of the environment was not the chairman’s intentional action, but *was* a known SE resulting from the chairman’s intentional action.

In the remaining samples (Question 1 for Sample 4.1 and Samples 4.5, and 4.6), with the exception of Sample 4.2 (where a marginal association was found), a majority of participants in each thought that questions or statements meant something other than was written, with  $\chi^2$ s ( $df = 1$  in Samples 4.1 [Question 1] and 4.6;  $df = 2$  in Sample 4.5,  $N$ s = 41, 43, and 42, respectively) ranging from 9.52 (Sample 4.6) to 23.77 (Sample 4.5) (all  $ps < .003$ ). For example, in Sample 4.1 (Question 1), most participants (82.9% across both conditions) did not think a statement about intentionally HH the environment referred to the chairman’s intentional action. Similarly, in Sample 4.6, most participants (73.8% across both conditions) thought the statement “In order to increase profits, the chairman intentionally HH the environment” actually meant “In order to increase profits, the chairman intentionally started a program he knew would HH the environment.”

These results provide additional strong support for our hypotheses, plainly showing that in half the samples and not dependent on

Table 6

*What Do the Statements and Questions “The Chairman Intentionally Helped/Harmed the Environment” and “Did the Chairman Intentionally Help/Harm the Environment?” Mean? Frequencies and Percentages of Selecting Each Response Option in Study 4*

Sample	Frequencies (%)	
	Help	Harm
Sample 4.1 ( $n = 41$ )/Question 1: Does “intentionally helping/harming” refer to chairman’s intentional action?		
Yes	2 (9.1)	5 (26.3)
No	20 (90.9)	14 (73.7)
Sample 4.1 ( $n = 41$ )/Question 2: Does the statement mean “the chairman intentionally started a program he knew would help/harm environment”?		
Yes	11 (50.0)	16 (84.2)
No	11 (50.0)	3 (15.8)
Sample 4.2 ( $n = 43$ ): Does the statement suggest the intentional action was to help/harm the environment, or that the helping/harming was a side effect of a different intentional action?		
Suggests that the intentional action was to help/harm the environment	12 (63.2)	9 (37.5)
Suggests that the helping/harming was a side effect of a different intentional action	7 (36.8)	15 (62.5)
Sample 4.3 ( $n = 45$ ): Is the question asking if the intentional action was to help/harm, or asking if the helping/harming was a known side effect of a different intentional action?		
It is asking if the intentional action was to help/harm	20 (83.3)	8 (38.1)
It is asking if the helping/harming was a known side effect of a different intentional action	4 (16.7)	13 (61.9)
Sample 4.4 ( $n = 42$ ): Is the question asking if the intentional action was to help/harm, or if the intentional action was to start a program that the chairman knew would help/harm?		
It is asking if the intentional action was to help/harm	18 (85.7)	10 (47.6)
It is asking if the intentional action was to start a program that he knew would help/harm	3 (14.3)	11 (52.4)
Sample 4.5 ( $n = 43$ ): Which of the following best captures the gist of the statement?		
He intentionally helped/harmed	2 (10.0)	2 (8.7)
He intentionally started a program to increase profits, knowing it would help/harm	13 (65.0)	16 (69.6)
He intentionally increased profits, knowing it would help/harm	5 (25.0)	5 (21.7)
Sample 4.6 ( $n = 42$ ): Does the statement mean he intentionally helped/harmed in order to increase profits, or that to increase profits, he intentionally started a program he knew would help/harm?		
Intentionally helped/harmed to increase profits	7 (35.0)	4 (19.2)
To increase profits, he intentionally started a program he knew would help/harm	13 (65.0)	18 (81.8)

*Note.* Significant chi-square associations between outcome condition and response option were found for Samples 4.1 (Question 2)–4.4 ( $ps \leq .02$ ) (for Sample 4.2, the association was marginally significant,  $p = .095$ ). For the remaining samples (4.1 [Question 1], 4.5, and 4.6), chi-square tests for the null hypothesis of equal proportions were significant,  $ps < .003$ .

outcome, majorities of participants thought that questions/statements about the intentionality of an SE outcome actually referred to an unmentioned intentional action (starting a program), and to the connection between that action and the SE outcome. This demonstrates that pragmatic understanding of statements and questions about intentionality—which differed from the literal meaning of the statements—was the same across both outcomes, involving an action that was inferred but not part of the questions/statements at all. Results in the other samples suggested that the way questions are interpreted can potentially account for differences in SE asymmetries. Presented with the helping versions, majorities took the questions at face value, suggesting that sometimes, participants who read about a morally beneficial SE outcome may disagree that the helping is intentional because they interpret the question as literally asking whether the helping is the chairman's intentional action, which it is not. Presented with the harm versions, though, participants read meaning into the question that is not written, reinterpreting it to focus on the intentional action that led to the harming. Thus, by giving different answers, participants provided with either outcome may be trying to draw similar conclusions—that the chairman did not intentionally help but that he did do something intentionally that led to harming. This suggests that, depending on how questions concerning intentionality of SE outcomes are asked, common conversational ground may not be as common as both researchers and research participants might think.

### Study 5

Study 5 was a replication study, using an undergraduate student sample and a new SE scenario, modified from Nadelhoffer (2006b). In it, a hunter in a bear-hunting contest shoots at a hungry-looking bear, knowing that hitting the bear will probably save an innocent bird-watcher from being attacked (beneficial outcome) or that the bullet will probably hit the innocent bird-watcher, seriously injuring him (harmful outcome). In both cases, the hunter does not care about the bird-watcher, just about winning the contest, and the bird-watcher is, respectively, harmed or helped. Participants answered a variety of questions used in earlier studies to see whether the same results would emerge using a new scenario in which the SE outcome involved a person (i.e., rather than an ambiguously defined “environment”), and in which the SE outcome was serious and less vague than the undefined and somewhat unclear HH of the environment.

### Method

Undergraduate participants ( $n = 65$ ) were randomly assigned to read either helping or harming versions of the bear scenario. Following this, they responded to five forced-choice questions, selecting whichever statement seemed most accurate from among the possible responses. The first choice was between “The hunter’s intention was to . . .” a) “HH the bird-watcher,” or b) “win the contest.” The second choice was between “The hunter . . .” a) “intentionally HH the bird-watcher,” or b) “did not intentionally HH the bird-watcher, but his beneficial (reckless) actions led to the bird-watcher being HH.” The third choice asked which of two statements best captured the meaning or gist of the statement “The hunter intentionally HH the bird-watcher.” The first response option referred back to the statement itself. The second option was

“By intentionally shooting the bear, the hunter beneficially helped (recklessly harmed) the bird-watcher.” The fourth choice was between “The hunter intentionally . . .” a) “HH the bird-watcher, in order to win the hunting contest,” and b) “shot the bear, in order to win the hunting contest, which led to the bird-watcher being HH.” The last choice was between “The hunter . . .” a) “intentionally HH the bird-watcher,” and b) “did not intentionally HH the bird-watcher. The benefit (harm) to the bird-watcher was an unintentional but known SE of the hunter’s actions.”

### Results and Discussion

Only one asymmetry was found across all questions (see Table 7). Specifically, a significant association was found between condition and response options regarding the hunter’s intentions,  $\chi^2(1, N = 65) = 4.13, p = .04$ . In this case, 12% (4/33) of the participants in the help condition indicated that the hunter’s intention was to help the bird-watcher, whereas all of the participants in the harm condition thought his intention was to win the hunting contest.

The remaining questions were examined using chi-square tests for the null hypothesis of equal frequencies of response across response options. Each of these tests was significant,  $\chi^2_s(1, N = 65)$  ranged from 11.22 (for the question related to the meaning of a statement about intentionality) to 46.54 (for the question related to intentionality vs. beneficial [reckless] action) (all  $ps \leq .001$ ). In each case, minorities in both groups thought the HH was intentional or intended. This strong support may have emerged because the HH was directed at a person rather than at a difficult-to-define environment. Furthermore, the HH was relatively extreme, vivid, and specific (i.e., saving a person from a bear attack vs. causing a person to be shot), whereas in the chairman vignette, helping or harming the environment is somewhat ambiguous.

Despite our finding here that few participants, when presented with other options, chose to say the chairman intentionally helped or harmed, we feel confident that if asked a single question such as “Did the hunter intentionally HH the bird-watcher?” the typical asymmetry would emerge. However, we again offered options that fully identified actions, goals, and SE outcomes, which allowed people to assign intentionality correctly to the action, and thus, only one (reverse) asymmetry was found. This replication, along with our earlier studies, provides compelling evidence that the badness of an SE outcome does not affect lay intuitions about intentionality in SE scenarios.

Yet, when asked particular types of questions, SE asymmetries consistently emerge. That is, even though people probably reinterpret questions about intentional HH to be asking about actions, and even though they see actions but not SE outcomes as intentional or not, the consistency of SE effects begs the question of why people are more willing to say an agent intentionally harmed than intentionally helped. Three brief final studies further examine these questions.

### Studies 6–8

The findings from Studies 3–5 show that when questions are properly framed, SE asymmetries disappear, and that when asked whether an agent intentionally brought about an SE outcome, perceivers, particularly in harm conditions, reframe the question to

Table 7

Replication of Earlier Effects Using a New (Bear-Hunting Contest) Scenario in Study 5 ( $n = 65$ )

Question and forced-choice response options (Option no.)	Frequency (%)	
	Help	Harm
Q1 (O1): The hunter's intention was to help/harm the bird-watcher	4 (12.1)	0 (00.0)
Q1 (O2): The hunter's intention was to win the contest	29 (87.9)	32 (100)
Q2 (O1): The hunter intentionally helped/harmed the bird-watcher	4 (12.1)	1 (3.1)
Q2 (O2): The hunter unintentionally but beneficially helped/recklessly harmed the bird-watcher	29 (87.9)	31 (96.9)
Q3 (O1): The statement means the hunter intentionally helped/harmed the bird-watcher	10 (30.3)	9 (28.1)
Q3 (O2): The statement means the hunter intentionally shot the bear, beneficially helping/recklessly harming the bird-watcher	23 (69.7)	23 (71.9)
Q4 (O1): The hunter intentionally helped/harmed the bird-watcher, in order to win the contest	3 (9.1)	4 (12.5)
Q4 (O2): The hunter intentionally shot the bear, in order to win the contest, leading to the helping/harming of the bird-watcher	30 (90.9)	28 (87.5)
Q5 (O1): The hunter intentionally helped/harmed the bird-watcher	7 (21.2)	4 (12.5)
Q5 (O2): The hunter did not intentionally help/harm the bird-watcher, but the helping/harming was a known side effect of the hunter's action	26 (78.8)	28 (87.5)

*Note.* O1 and O2 refer to possible response options for each question. Questions asked participants to pick the statement that best described details from the scenario, or asked about the meaning of the statement "The hunter intentionally helped/harmed the bird-watcher." For Question (Q) 1, there was a significant chi-square association between condition and response option ( $p = .04$ ). For the remaining questions, chi-square tests for the null hypothesis of equal frequencies across response options were significant,  $ps \leq .001$ .

focus on the agent's action. By showing that people prefer to say the chairman knowingly, rather than intentionally, harmed, Guglielmo and Malle (2010a) provided one clue as to why, despite this, perceivers are more likely to attribute intentionality to harmful versus beneficial SE outcomes. Beliefs (i.e., an agent's knowledge connecting actions to outcomes) are foundational to intentionality judgments because an action cannot be intentional when an acting agent does not perceive its connection to an outcome. However, without an intention, there can also be no intentional action. When knowledge and intention are aligned in their connection to a single outcome (i.e., an agent knows an action will lead to an intended outcome), the action is clearly intentional as long as other components of intentionality are present. However, when evaluating SE scenarios, perceivers are confronted with an agent who unambiguously intentionally acts, knowing the action will lead not only to an intended outcome but also to another (SE) outcome that is not specifically intended. Thus, the agent's knowledge and intention regarding the SE outcome are misaligned, and perceivers are faced with a choice. When asked about the intentionality of the second, unintended outcome, they must decide which mental state is more important—the agent's knowledge or intentions.

The importance of intentions/goals (i.e., motives) to understanding intentional action and, more generally, to understanding how people make trait attributions was discussed by Reeder (2009a, 2009b) in his elaboration of a multiple inference model (MIM) of person perception. In brief, Reeder suggests that on the way to attributing traits, perceivers will create narratives about agents' intentional behaviors to help explain why they acted as they did (see also Trafimow, 2009). Naturally, inferring an agent's motives—one type of mental state—can help perceivers understand what "type" of person an agent is, because situational constraints aside, motives for good and bad behavior should inform character judgments. However, an already complex attribution process may become even more complex when an agent's behavior leads not only to an intended outcome but also to a secondary outcome about which the agent knows but does not care. That is, an agent's motives may

become more or less important for trait attribution relative to other mental states such as an agent's knowledge or awareness about a side effect of his or her action, depending on factors such as the moral valence of the SE and whether the original motive was prosocial, antisocial, or morally neutral. Although the current studies were not designed to answer this question in particular, we think they might provide additional insight into how people create sensible stories about others' behavior (e.g., Pennington & Hastie, 1992).

Specifically, we believe that perceivers attend to different mental states when harmful versus beneficial SEs are brought about and that this attention serves a social function. As others have noted (e.g., Bartsch & Young, 2010; Lombrozo & Uttich, 2010; Malle & Nelson, 2003; Uttich & Lombrozo, 2010; Ybarra, 2002), there are important reasons for perceivers to attend to bad, norm-violating behavior, even while they ignore or dismiss good, norm-conforming behavior. When presented with the helping version of the chairman vignette, perceivers probably focus on the chairman's intention/goal, and what he did in service of this goal, to decide whether he deserves credit for his action. They might wonder, "Was he *intending* to cause benefit? Was benefit a *goal* of his action?" Answering no, they dismiss his knowledge as less socially meaningful, and say he did not intentionally help because he did not intend to help and the helping was not his intentional action. In the harm case, however, perceivers become "outcome focused," because when something bad happens, people want to know why it happened, and who was responsible. Knowledge becomes socially meaningful, and perceivers might wonder, "Did he *know* that his (intentional) action would lead to the outcome? Was he aware, while acting, that his action would harm?" Answering yes, they dismiss his morally neutral motive for action as less socially meaningful, because even if harm was not intended, he knew it would occur. In this case, knowledge rather than motive might contribute more to a character judgment.

If this hypothesis is correct, when presented with different SE outcomes and asked about the intentionality of these outcomes,

participants' decisions about intentionality should place greater weight on the importance of intentions/goals relative to knowledge/awareness when a helpful SE outcome comes about, with the reverse true when a harmful outcome occurs. Decisions should also be more focused on his action relative to the SE outcome in the helpful case, with the reverse true in the harmful case. Furthermore, if these shifts in focus serve a socially useful function, participants should think the chairman's action will be more predictive of his future behavior in the harmful case relative to the helpful one.

## Method

**Study 6.** Participants ( $n = 96$ ) were randomly assigned to read either a helping or a harming version of a brief, abstract paragraph about "John," who does something neutral to achieve a self-serving goal, knowing that something else good/bad (about which he does not care) might happen as a result of his action. Participants were then asked four questions. Question 1 asked for a choice between whether John intentionally (a) "did something to achieve a self-serving goal, which incidentally brought about a beneficial (harmful) outcome," or (b) "brought about a beneficial (harmful) outcome, in order to achieve a self-serving goal." Participants were then asked what "intentionally" referred to on a 10-point scale (1 = *John's action*, 10 = *The good/bad outcome that resulted from his action*) and rated their agreement (1 = *completely disagree*, 10 = *completely agree*), in two separate questions, that when a beneficial (harmful) outcome comes about, it is important to know why it happened, and whose actions caused it.

**Studies 7–8.** In both of these studies, participants read either helping or harming versions of the chairman vignette and rated their agreement (using the same 10-point agreement scale as above) with the statement "The chairman intentionally HH the environment." Following this, participants in Study 7 ( $n = 95$ ) answered three questions about what most influenced their decisions regarding intentionality, with response options contrasting the chairman's intentions against his knowledge of HH (1 = *his intentions*, 10 = *his knowledge*), his goal of increasing profits against his awareness that his actions would lead to HH (1 = *his goal*, 10 = *his awareness*), and his action (i.e., starting a program) against the HH outcome that resulted from it (1 = *his action*, 10 = *the HH outcome*). Because the knowledge and awareness questions were strongly correlated ( $r = .74$ ), these two items were averaged to create a composite variable contrasting the chairman's intentions/goals against his knowledge/awareness. In Study 8 ( $n = 60$ ), two separate questions using 10-point scales (1 = *not at all important*, 10 = *very important*) asked about the importance of the chairman's intentions and knowledge for their understanding of intentionality in the statement with which they rated their agreement (i.e., "The chairman intentionally HH the environment"). To index the relative weight of intentions versus knowledge, responses to the intention question were subtracted from those to the knowledge question, so that positive values would indicate a greater weight on knowledge relative to intentions, and negative values the reverse. A final question asked about the extent to which the chairman's (current) behavior would be informative regarding his future behavior (1 = *not much*, 10 = *a lot*).

## Results

**Study 6.** A majority of people across both conditions selected the definition that emphasized incidental HH resulting from the chairman's intentional action (82/96; 85.4%) rather than his intentionally bringing about the good/bad outcome,  $\chi^2(1, N = 96) = 48.17, p < .001$ . However, relative to people in the help condition (3/49; 6.1%), a greater proportion of people in the harm condition (11/47; 23.4%) selected that the agent intentionally brought about the outcome,  $\chi^2(1, N = 96) = 5.75, p = .02$ . In the harm condition ( $M = 4.36, SD = 3.31$ ) relative to the help condition ( $M = 2.88, SD = 2.81$ ), participants reported that their responses were influenced more by the outcome versus the chairman's action,  $t(94) = 2.37, p = .02, d = .48$ . Likewise, in the harm condition, people thought it was more important to know why the outcome occurred ( $M = 7.94, SD = 1.87$ ) and whose actions caused it ( $M = 7.77, SD = 2.09$ ), relative to the help condition ("why"  $M = 6.47, SD = 1.72$ ; "caused"  $M = 6.80, SD = 1.98$ ; respectively,  $t_s[94] = 4.00$  and  $2.24, p_s = .0001, .02, d_s = .82, .48$ ). Point-biserial correlations of definition (0 = incidental HH, 1 = intentional HH) with action/outcome ( $r = .38$ ), why ( $r = .22$ ), and who caused ( $r = .22$ ) questions were significant ( $p_s < .05$ ). A greater focus on outcomes was not correlated with importance of knowing why it happened ( $r = .12, ns$ ) or who caused it ( $r = .16, ns$ ), but why and cause were significantly correlated ( $r = .68, p < .001$ ).

**Study 7.** HH condition significantly affected responses to all dependent variables. Relative to the help condition ( $M = 2.35, SD = 2.26$ ), participants in the harm condition ( $M = 8.59, SD = 1.99$ ) agreed more that the chairman intentionally harmed the environment,  $t(93) = 14.31, p < .001, d = 2.93$ . However, in arriving at their decisions regarding intentionality and relative to the help condition, participants in the harm condition reported placing greater weight on his knowledge/awareness relative to his intention/goal (harm  $M = 7.27, SD = 1.71$ ; help  $M = 2.40, SD = 1.78$ ),  $t(93) = 13.60, p < .001, d = 2.79$ , and greater weight on the outcome relative to his action (harm  $M = 5.39, SD = 2.88$ ; help  $M = 2.74, SD = 2.27$ ),  $t(93) = 4.96, p < .001, d = 1.02$ . Intentionality ratings were strongly correlated with greater focus on knowledge/awareness relative to intentions/goals ( $r = .86, p < .001$ ), and also with a greater focus on outcomes relative to actions ( $r = .44, p < .001$ ). A focus on outcomes was correlated with a focus on knowledge ( $r = .53, p < .001$ ).

**Study 8.** Condition again significantly predicted all variables. Participants agreed more that the chairman intentionally harmed ( $M = 8.34, SD = 2.07$ ) than intentionally helped ( $M = 2.22, SD = 1.94$ ),  $t(58) = 11.79, p < .001, d = 3.05$ . Participants in the harm condition also placed a greater weight on knowledge relative to intentions ( $M = 1.55, SD = 4.31$ ) compared with those in the help

<sup>11</sup> Independent analyses of these variables (i.e., knowledge and intentions) showed that in the help condition (intentions  $M = 8.29, SD = 2.22$ ; knowledge  $M = 5.16, SD = 3.00$ ) relative to the harm condition (intentions  $M = 6.48, SD = 3.34$ ; knowledge  $M = 8.03, SD = 2.38$ ), there was a greater focus on the chairman's intentions,  $t(58) = 2.48, p = .02, d = .64$ , and a lesser focus on his knowledge,  $t(58) = 4.09, p < .001, d = 1.06$ . Furthermore, greater focus on knowledge was associated with stronger agreement about intentionality ( $r = .54, p < .001$ ), whereas greater focus on intentions was associated with weaker agreement about intentionality ( $r = -.32, p = .01$ ). Focus on intentions was not significantly correlated with focus on knowledge ( $r = -.13, p = .32$ ).

condition ( $M = -3.13$ ,  $SD = 3.42$ ),  $t(58) = 4.67$ ,  $p < .001$ ,  $d = 1.20$ ,<sup>11</sup> and thought the chairman's behavior would be more predictive of his future actions (harm  $M = 8.66$ ,  $SD = 1.84$ ; help  $M = 6.06$ ,  $SD = 2.85$ ),  $t(58) = 4.15$ ,  $p < .001$ ,  $d = 1.08$ . Intentionality was correlated with greater focus on knowledge relative to intentions ( $r = .57$ ) and prediction of the chairman's future actions ( $r = .55$ ;  $ps < .001$ ). Greater focus on knowledge relative to intentions was correlated with stronger prediction of future actions ( $r = .26$ ,  $p = .04$ ).

## Discussion

These final three studies support the hypotheses that considering harmful versus helpful outcomes influences the extent to which perceivers focus on an agent's action versus an SE outcome that resulted from the action. It also influences the mental states perceivers feel are important in making their decisions about intentionality, focusing them to a greater extent on the agent's knowledge relative to his intentions when the SE outcome is harmful, with the reverse true when the SE outcome is helpful. Study 6 also provided suggestive evidence that the reason for these shifts in focus is the greater importance placed on knowing why harmful, relative to beneficial, outcomes occur, and who caused them. Finally, participants presented with a harmful outcome relative to a helpful one thought the chairman's action was more predictive of how he might act in the future, showing that differences in intentionality attributions may be serving a useful social purpose of behavioral prediction.

Although these studies were not conducted specifically to do so, they complement Reeder's (2009a, 2009b) MIM and have the potential to spark further theoretical development in extending the MIM to situations when an agent's behavior leads not only to an intended outcome but also to a known (but not specifically intended) secondary outcome. In short, the MIM suggests that people's inferences about agents' motives for acting intentionally provide one path to inference of more stable traits or character judgments. The current studies suggest that this process may become more complex when considering unintended secondary outcomes caused by the agent's actions and known to the agent. For example, highlighting the greater importance of knowledge over motive in some cases, it is easy to imagine how even the most prosocial motive and outcome could lead to a negative rather than positive character evaluation if an agent knows (and does not care that) his or her action will lead to a good person being secondarily harmed by an action. Similarly, negative trait ascriptions could be exacerbated when a primary antisocial motive and outcome is followed by a secondary anticipated and harmful outcome. More generally, when the SE outcome involves a second agent, evaluations of *that* agent's character might also play a role in character evaluations of the primary, acting agent (e.g., when an agent does a good thing for a good person, knowing that a bad person will also be harmed). Future work should consider these and similar issues, which suggest interesting predictions about how people make trait judgments in increasingly complex situations—similar to how people do so regularly in everyday life.

## General Discussion

In this article, our goal was to show that although SE effects reliably emerge when people are asked about the intentionality of

SE outcomes (e.g., Knobe, 2010b), SE outcomes themselves do not influence intuitions about intentional action. Our argument began with the idea that intentionality is a complex construct requiring inferences about multiple mental states (e.g., Malle & Knobe, 1997a) and their relationship to actions, goals, and outcomes. In conjunction with this, we argued that actions and outcomes are distinct and that lay understanding of the word *intention* often reflects a focus on outcomes and goals, rather than on actions. However, when people think about intentionality, they are typically considering *intentional actions*, not goals or outcomes, and particularly not secondary outcomes that were not specifically intended. We offered this interpretation because people can control (at least to some extent) their actions and goals, but not whether their actions will result in desired and intended outcomes. Study 1 supported these hypotheses, demonstrating that people differentiate between actions (things people *do*), outcomes (things that *happen*, often as a result of things people do), goals (things people want to *accomplish*, or ends toward which effort is directed), and SE (*unintended outcomes* that may or may not be foreseen). We also showed in several ways that most people associate intentions with outcomes and goals, but associate intentionality with the things people do, helping to resolve the question of "intentionality without intention" (e.g., Knobe, 2003b). In addition, we showed that people agree in their labeling of each of these components in an SE scenario, even when intentionality is not mentioned.

Building on this, we argued that when questions ask whether a person intentionally did something, people expect the "did" to be an action, and noted that questions about intentional helping and harming linguistically turn SE outcomes into actions. We also considered how SE results are typically discussed in terms of intentional actions (e.g., Knobe, 2003a), suggesting the possibility of a mismatch between participants' pragmatic understanding of questions about intentional SE and the way their responses are semantically interpreted by people studying this effect. For example, a person might say, "John (intentionally) helped his friend," which makes it sound as if the helping was John's action. Nevertheless, this sentence implies a different action that led to helping, such as "[By doing X.] John helped," and it is this action that is intentional, not the outcome. However, it is important to note that in this simple example, the agent's knowledge connects his motivated action (X) to a goal/intended outcome (to be helpful), which is in turn served by the agent's action. In SE scenarios, questions about intentionality become more complicated. This is because the agent's knowledge connects his action to more than one outcome, only one of which is intended. The other outcome is unintended but, because of the agent's beliefs (i.e., knowledge), foreseen. In these cases, if all actions, goals, and outcomes are properly identified in response options, helpful and harmful outcomes should lead to the same answers about intentionality. That is, most people should prefer to say that the agent intentionally *acted*, which *led* to the SE outcome, rather than saying the agent intentionally *brought about* the outcome. Related to this, we suspected that people would reinterpret the meaning of statements or questions about "intentionally HH" as asking whether agents intentionally acted, knowing the SE outcome would occur, because the reinterpretation including action makes more sense than a literal one that asks whether an agent intentionally (or even knowingly) brought about an outcome.

In Study 3, we demonstrated that if all actions, outcomes, and mental states of the agent (i.e., intentions/goals and knowledge) are properly identified and offered as response options, almost no participants responded that the agent intentionally HH (Samples 3.1 and 3.2); instead, people selected responses that focused on the chairman's action (starting a program) in service of a goal, even when they could have called both this action *and* the outcome intentional. As expected, no SE asymmetries for intentionality were found. Furthermore, when rating each response option separately (Samples 3.3 and 3.4), participants agreed that it sounded right to say he intentionally acted, and wrong to say he intentionally HH (or any other option). In Study 4, across six samples, we further confirmed our hypotheses by showing that when people are presented with statements and questions about intentional HH, majorities think the statements/questions are actually focusing on the agents' intentional action and the link between the action and the action's consequences. Providing further strong support, in half the samples, HH asymmetries were found, showing that although a majority of participants in the harm conditions thought the questions/statements were referring to unmentioned (and intentional) *actions*, a majority of participants in the help conditions thought the questions were literally asking whether the chairman intentionally brought about an outcome. This suggests that when outcomes are helpful versus harmful, different ways of understanding the same questions may also drive SE effects. Together, Studies 3 and 4 definitively showed that outcomes do not really seem to affect lay intuitions about intentionality—at least in the cases of SE scenarios—which appear to typically refer to actions, and only extremely rarely to SE outcomes. These studies also suggest that to understand the role of intentional action in bringing about SEs, it is important to provide response options that do not oversimplify the relations between actions, mental states, and outcomes, because when simplified questions *are* asked, people may not be answering the questions one is intending to ask.

We further confirmed these conclusions in Study 5, where we replicated findings from earlier studies using a new scenario. In fact, this study suggested that when people are given scenarios containing more concrete outcomes (e.g., injuring a person vs. saving a person from injury) rather than ambiguous ones (e.g., harming vs. helping the environment), people are even more reluctant to assign intentionality to SE outcomes when other options are available.

Finally, we argued that harmful and beneficial outcomes are structurally different, in part because actions leading to harmful outcomes should be avoided, whereas actions leading to beneficial outcomes do not necessarily need to be encouraged. Study 2 confirmed this, showing that statements about allowing (causing, being responsible for) help and harm are not seen in the same way. Saying someone allowed harm sounds correct, but saying someone allowed help sounds neither right nor wrong, at best. In addition, Study 2 demonstrated that people think it is more important to avoid harm than to promote benefit.

This makes sense, because it serves a social purpose to be concerned when harm is knowingly brought about (e.g., Bartsch & Young, 2010). We therefore hypothesized that when faced with a harmful outcome relative to a helpful outcome, people would have a greater desire to know why it occurred and who caused it. Also, because of differences in how people think about harming and helping, we expected that harmful outcomes would shift attention

away from agents' actions and intentions and focus it on the harmful outcomes themselves, including the agents' knowledge that the outcomes would result from their actions. Ultimately, these shifts in attention should serve a social function, because knowing that a person will choose to act, even when this action will likely bring about a foreseen negative consequence, tells an observer something about how this agent will act in the future. Helpful outcomes, however, contain less socially valuable information. Acting in a way that brings about unintended but foreseen benefit does not tell a perceiver much about how the agent will act in the future, because acting good is the normative default (e.g., Guglielmo & Malle, 2010a; Uttich & Lombrozo, 2010; Ybarra, 2002). In fact, when an agent knows his actions will cause benefit, but does not seem to care, it makes clear that the benefit is not intended. Perceivers then turn to the agent's *motives* for acting (i.e., his intentions or goals), and weight this information more heavily than his knowledge.

In our final studies (6–8), we provided evidence that good and bad outcomes influence *which* mental states people think are important for making intentionality judgments. That is, people presented with a helpful outcome focused on an agent's intentions (i.e., the agent's motive for acting), whereas harmful outcomes focused people on an agent's beliefs (the agent's knowledge that an action will lead to an outcome). Along with this, helping focused people on the action the chairman took to further his intention (starting a program), whereas harming focused people on the known SE outcome that resulted from his action. This attention to different mental states and components of the action/outcome sequence seems to serve the purpose of trying to understand how the agent will act in the future. Potentially, it also helps perceivers in making character judgments about the agent. That is, although focused on questions about intentional actions and SE, findings from Studies 6–8 offer the potential for new theorizing and empirical work on the related topic of trait ascription in special cases.

## Limitations and Future Directions

Although we believe we have in large part addressed the question of why SE asymmetries emerge and shown that good versus bad SE do not influence intuitions about intentionality, we note a few potential limitations. First, a number of our studies were conducted using a single scenario. However, this scenario has been used in many studies to demonstrate SE effects, and has been widely discussed, making it an ideal candidate for use here. Furthermore, we confirmed our findings from other studies using a different scenario in Study 5, and results using this scenario were even more consistent with our arguments than in other studies, suggesting that SE effects may most easily emerge when SE outcomes are ambiguous, as is the entity that is helped or harmed. Second, we primarily focused on intentions and intentionality. Although true, we also investigated the effect of SE outcomes on a variety of other mental state variables (e.g., knowledge, awareness). Furthermore, intentionality is perhaps the most studied outcome in connection with SE effects (see Knobe, 2010b). By showing that intuitions about intentionality are not affected by SE outcomes, we cast doubt on the extent to which SE outcomes affect other variables of interest.

In Studies 7 and 8, although the variables we included explained substantial variance in intentionality judgments (i.e., between 37% and 76%), some of the variability in these judgments remained unaccounted for. However, fully explaining all variance in intentionality judgments might require concurrently asking about people's understanding of the questions, the agent's mental states, their focus on outcomes versus actions, and the "purpose" of the questions, while also taking into account structural differences between HH and the social functions of blame and praise. We expect that it might be difficult to simultaneously disentangle the complex relationships among all these variables in understanding their contribution to attributing intentionality. Furthermore, in these two studies, we did not ask whether it "sounds right or wrong" to say the chairman intentionally HH (a response scale that is often used); we asked for agreement that he did intentionally HH. Given that prior research has shown it typically does not sound very right to say he intentionally harmed, if we had used this question, there probably would have been substantially less variability in responses as a function of HH to explain.

Causality among these variables might be difficult to establish, unless each putative explanatory variable is systematically manipulated. Intentionality judgments are also made very quickly (e.g., Malle & Holbrook, 2012), and when studying intentionality in relation to SEs, people may quickly make snap judgments as to the intentionality of outcomes, only realizing later (if ever) that what they *meant* was that the agent's action was intentional and caused the outcome. Questions such as these can probably best be answered by testing whether, in the harm case, people respond more quickly to questions about the agent's beliefs than to questions of intentional harming, and in the help case, respond more quickly to questions about intention. It might also be useful to investigate whether it takes longer to respond to helping or harming outcomes. One hypothesis is that it should take longer to determine intentionality in the harming case because people first need to disentangle that the question being asked is about the agent's intentional action in service of a goal, while also attending to the fact that he knew a harmful SE would come about, and that he did not intend it.

In the end and at the least, we offer a cautionary note that future research on this effect should proceed carefully and in full consideration of the issues raised here. Moreover, we offer a simple suggestion that might help encourage a conversational common ground between researchers and their participants when conducting future research on the SE effect: Make sure that there is common understanding of the meaning of statements and questions about intentionality and other mental states, perhaps by simply asking participants to describe what these statements and questions pragmatically mean, in their own words.

## Conclusion

Over the last 10 years, a large quantity of research has suggested that moral judgments influence lay intuitions about intentionality (e.g., Knobe, 2010b). This consequential idea contradicts the long-held view that intentionality ascriptions affect moral judgments, and the moral influence view has been challenged on a number of grounds (e.g., Adams & Steadman, 2004a, 2004b; Guglielmo & Malle, 2010a, 2010b; Guglielmo et al., 2009; McGuire, 2012; Uttich & Lombrozo, 2010; Wiland, 2007), perhaps because of the

real-world consequences if this hypothesis is correct. The current article offers an alternative view that complements and extends this prior work. One contribution of the current research is to point out that calling an SE an action (e.g., Cushman & Mele, 2008; Knobe, 2003a; McGuire, 2012) is incorrect, because SEs, as a class, are unintended outcomes that result from actions, whether foreseen or not. Thus, questions about the intentionality of SEs (and outcomes, more generally) are misleading. When questions properly identify actions along with goals and outcomes, people typically agree that it is the action—which remains constant across helpful and harmful outcomes—that is intentional. In fact, people understand questions about intentionally caused outcomes to be questions about intentional actions that lead to those outcomes. Thus, our findings refute the idea that outcomes influence intuitions about the intentionality of SE. In the end, we make a somewhat modest observation: Responses to simple questions may not provide trustworthy answers to complex questions, because, in short, questions matter.

## References

- Adams, F. (1986). Intention and intentional action: The simple view. *Mind and Language*, 1, 281–301. doi:10.1111/j.1468-0017.1986.tb00327.x
- Adams, F., & Steadman, A. (2004a). Intentional action and moral considerations: Still pragmatic. *Analysis*, 64, 268–276. doi:10.1093/analys/64.3.268
- Adams, F., & Steadman, A. (2004b). Intentional action in ordinary language: Core concept or pragmatic understanding? *Analysis*, 64, 173–181. doi:10.1093/analys/64.2.173
- Alvarez, M. (2009). Acting intentionally and acting for a reason. *Inquiry*, 52, 293–305. doi:10.1080/00201740902917168
- Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences*, 5, 171–178.
- Bartsch, K., & Young, T. (2010). Reasoning asymmetries do not invalidate theory-theory. *Behavioral and Brain Sciences*, 33, 331–332. doi:10.1017/S0140525X10001688
- Cushman, F., Knobe, J., & Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition*, 108, 281–289. doi:10.1016/j.cognition.2008.02.005
- Cushman, F., & Mele, A. (2008). Two-and-a-half folk concepts? In J. Knobe & S. Nichols (Eds.), *Experimental philosophy* (pp. 171–188). Oxford, England: Oxford University Press.
- Gintis, H. (2010). Modalities of word usage in intentionality and causality. *Behavioral and Brain Sciences*, 33, 336–337. doi:10.1017/S0140525X10001731
- Grice, H. P. (1989a). Further notes on logic and conversation. In H. P. Grice (Ed.), *Studies in the way of words* (pp. 41–57). Cambridge, MA: Harvard University Press. (Reprinted from *Syntax and semantics*, 9: *Pragmatics*, pp. 113–128, by P. Cole, Ed., New York, NY: Academic Press)
- Grice, H. P. (1989b). Logic and conversation. In H. P. Grice (Ed.), *Studies in the way of words* (pp. 22–40). Cambridge, MA: Harvard University Press. (Reprinted from *Syntax and semantics*, 3: *Speech acts*, pp. 41–58, by P. Cole & J. Morgan, Eds., 1975, New York, NY: Academic Press)
- Guglielmo, S., & Malle, B. F. (2010a). 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
- Guglielmo, S., & Malle, B. F. (2010b). Enough skill to kill: Intentionality judgments and the moral valence of action. *Cognition*, 117, 139–150. doi:10.1016/j.cognition.2010.08.002

- Guglielmo, S., Monroe, A. E., & Malle, B. F. (2009). At the heart of morality lies folk psychology. *Inquiry*, *52*, 449–466. doi:10.1080/00201740903302600
- Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley. doi:10.1037/10628-000
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 371–388). Hillsdale, NJ: Erlbaum.
- Knobe, J. (2003a). Intentional action and side effects in ordinary language. *Analysis*, *63*, 190–194. doi:10.1093/analys/63.3.190
- Knobe, J. (2003b). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, *16*, 309–324. doi:10.1080/09515080307771
- Knobe, J. (2004a). Folk psychology and folk morality: Response to critics. *Journal of Theoretical and Philosophical Psychology*, *24*, 270–279. doi:10.1037/h0091246
- Knobe, J. (2004b). Intention, intentional action and moral considerations. *Analysis*, *64*, 181–187. doi:10.1093/analys/64.2.181
- Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies*, *130*, 203–231. doi:10.1007/s11098-004-4510-0
- Knobe, J. (2007). Reason explanation in folk psychology. *Midwest Studies in Philosophy*, *31*, 90–106. doi:10.1111/j.1475-4975.2007.00146.x
- Knobe, J. (2010a). The person as moralist account and its alternatives. *Behavioral and Brain Sciences*, *33*, 353–365. doi:10.1017/S0140525X1000230X
- Knobe, J. (2010b). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, *33*, 315–329. doi:10.1017/S0140525X10000907
- Knobe, J., & Burra, A. (2006). The folk concepts of intention and intentional action: A cross-cultural study. *Journal of Cognition and Culture*, *6*, 113–132. doi:10.1163/156853706776931222
- Levinson, S. C. (1983). *Pragmatics*. Cambridge, England: Cambridge University Press.
- Lombrozo, T., & Uttich, K. (2010). Putting normativity in its proper place. *Behavioral and Brain Sciences*, *33*, 344–345. doi:10.1017/S0140525X10001810
- Malle, B. F. (2006). Intentionality, morality, and their relationship in human judgment. *Journal of Cognition and Culture*, *6*, 87–112. doi:10.1163/156853706776931358
- 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
- Malle, B. F., & Knobe, J. (1997a). The folk concept of intentionality. *Journal of Experimental Social Psychology*, *33*, 101–121. doi:10.1006/jesp.1996.1314
- Malle, B. F., & Knobe, J. (1997b). Which behaviors do people explain? A basic actor–observer asymmetry. *Journal of Personality and Social Psychology*, *72*, 288–304. doi:10.1037/0022-3514.72.2.288
- Malle, B. F., & Knobe, J. (2001). The distinction between desire and intention: A folk-conceptual analysis. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 45–67). Cambridge, MA: MIT Press.
- Malle, B. F., & Nelson, S. E. (2003). Judging mens rea: The tension between folk concepts and legal concepts of intentionality. *Behavioral Sciences and the Law*, *21*, 563–580. doi:10.1002/bsl.554
- McGuire, J. M. (2012). Side-effect actions, acting for a reason, and acting intentionally. *Philosophical Explorations*, *15*, 317–333. doi:10.1080/13869795.2012.696130
- Nadelhoffer, T. (2006a). Bad acts, blameworthy agents, and intentional actions. *Philosophical Explorations*, *9*, 203–220. doi:10.1080/13869790600641905
- Nadelhoffer, T. (2006b). Desire, foresight, intentions, and intentional actions: Probing folk intuitions. *Journal of Cognition and Culture*, *6*, 133–157. *Philosophical Explorations*, *9*, 203–220. doi:10.1080/13869790600641905
- Núñez, N. L., Laurent, S. M., & Gray, J. M. (2014). Is negligence a first cousin to intentionality? Lay conceptions of negligence and its relationship to intentionality. *Applied Cognitive Psychology*, *28*, 55–65. doi:10.1002/acp.2957
- Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology*, *62*, 189–206. doi:10.1037/0022-3514.62.2.189
- Pettit, D., & Knobe, J. (2009). The pervasive impact of moral judgment. *Mind and Language*, *24*, 586–604. doi:10.1111/j.1468-0017.2009.01375.x
- Reeder, G. D. (2009a). Mindreading and dispositional inference: MIM revised and extended. *Psychological Inquiry*, *20*, 73–83. doi:10.1080/10478400902868565
- Reeder, G. D. (2009b). Mindreading: Judgments about intentionality and motives in dispositional inference. *Psychological Inquiry*, *20*, 1–18. doi:10.1080/10478400802615744
- Scaife, R., & Webber, J. (2013). Intentional side-effects of action. *Journal of Moral Psychology*, *10*, 179–203. doi:10.1163/17455243-4681004
- Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Hillsdale, NJ: Erlbaum.
- Shaver, K. G. (1985). *The attribution of blame*. New York, NY: Springer-Verlag. doi:10.1007/978-1-4612-5094-4
- Trafitimow, D. (2009). Reeder's MIM as a special case of confluence theory. *Psychological Inquiry*, *20*, 48–52. doi:10.1080/10478400902744311
- Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition*, *116*, 87–100. doi:10.1016/j.cognition.2010.04.003
- Vallacher, R. R., & Wegner, D. M. (2012). Action identification theory: The highs and lows of personal agency. In P. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories in social psychology* (pp. 327–348). London, England: Sage.
- Wiland, E. (2007). Intentional action and “in order to.” *Journal of Theoretical and Philosophical Psychology*, *27*, 113–118. doi:10.1037/h0091285
- Ybarra, O. (2002). Naïve causal understanding of valenced behaviors and its implications for social information processing. *Psychological Bulletin*, *128*, 421–441. doi:10.1037/0033-2909.128.3.421

Received January 29, 2014

Revision received May 2, 2014

Accepted May 5, 2014 ■