

The Moral Dimension of Intentionality Judgments

Bertram F. Malle

*Institute of Cognitive and Decision Sciences
and Department of Psychology, University of Oregon*

Joshua Knobe (2003a, b) presented data that suggest people's judgments of a behavior's intentionality may be significantly influenced by moral considerations. In particular, Knobe (2003b) argues that when people judge the intentionality of an action with moral consequences, they fail to consider an important component of intentionality — the agent's skill (Malle & Knobe, 1997) — that they do consider when judging an almost identical action with neutral consequences. This finding raises a number of issues about the consistency of intentionality judgments and perhaps even the unity of the folk concept of intentionality. Moreover, it raises the specter of a bias in people's thinking, namely to ignore important information when judging morally significant actions, which, if true, would have considerable impact on legal proceedings.

A replication of Knobe's results appears to be needed, not only to establish the findings within a different subject population (in this case, first- and second-year college students) but also to examine the robustness of Knobe's effect when varying certain methodological and conceptual parameters. The parameters that were varied in the present study ranged from presentation format to action formulation to response options.

Just over 300 participants were presented with a one-page questionnaire as part of a larger survey packet. The sample consisted of 74% women, 77% Caucasians, with a median age of 18 years. The questionnaire asked participants to respond to Knobe's (2003b) rifle scenario, which was either an act of shooting a bull's eye with the morally neutral consequence of winning a contest or an act of shooting a person with the immoral consequence of killing the person in an attempt to acquire inheritance money.

The scenarios were as follows:

Contest

Jake desperately wants to win the rifle contest. He knows that he will only win the contest if he hits the bulls-eye. He raises the rifle, gets the bull's-eye in the sights, and pulls the trigger. Jake is an expert marksman. His hands are steady. The gun is aimed perfectly... The bullet lands directly on the bull's-eye. [*Luck*: But Jake isn't very good at using his rifle. His hand slips on the barrel of the gun, and the shot goes wild, bouncing off a heavy post. Nonetheless, the bullet lands directly on the bull's-eye.] Jake wins the contest.

Inheritance

Jake desperately wants to have more money. He knows that he will inherit a lot of money when his aunt dies. One day, he sees his aunt walking by the window. He raises his rifle, gets her in the sights, and pulls the trigger. Jake is an expert marksman. His hands are steady. The gun is aimed perfectly... The bullet hits her directly in the heart. [*Luck*: But Jake isn't very good at using his rifle. His hand slips on the barrel of the gun, and the shot goes wild, bouncing off a heavy post. Nonetheless, the bullet hits her directly in the heart.] She dies instantly.

The report of the findings of this study are organized into five sections: Overall replication; effects of presentation format; effects of action verb choice; and effects of response options.

1. Overall Replication

Of the entire sample of 228 participants, 155 were presented with the neutral (contest) scenario, 73 with the moral (inheritance) scenario, and of each group, half saw the *Luck* version and half saw the *Skill* version, resulting in a 2 (moral vs. neutral) \times 2 (luck vs. skill) design. The first research question was whether we could replicate Knobe's basic results. Table 1 shows people's blame and praise ratings and Table 2 shows people's intentionality ratings.

Table 1. Evaluations (praise or blame) of stimulus action under conditions of luck or skill.

	<i>Luck</i>	<i>Skill</i>
How much praise does Jake deserve for hitting the bull's eye? [from 0, no praise, to 6, a lot of praise]	3.1	5.2
How much blame does Jake deserve for shooting* his aunt? [from 0, no blame, to 6, a lot of blame]	5.8	5.8

* Note: The displayed means are averaged across three verb forms (hit, shoot, kill).

Table 2. Intentionality judgments of stimulus action under conditions of luck or skill.

	<i>Luck</i>	<i>Skill</i>
Did Jake hit the bull's-eye intentionally ? (Yes, No)	30%	90%
Did Jake shoot* his aunt intentionally ? (Yes, No)	85%	97%

* Note: The displayed means are averaged across three verb forms (hit, shoot, kill).

The results displayed in Tables 1 and 2 clearly replicate Knobe's (2003b) findings, such that people differentiate between hitting a bull's eye with luck and hitting it with skill but do not sufficiently differentiate the parallel case of shooting a person with luck or shooting her with skill. This differentiation in the neutral scenario is substantial for the evaluative ratings, $F(154) = 33.6, p < .001, \eta^2 = 18\%$, as well as for the intentionality judgments, $F(154) = 30.3, p < .001, \eta^2 = 16\%$. By contrast, in the moral scenario no differentiation exists for evaluative ratings, $F < 1, \eta^2 = 0\%$, and only a marginally significant differentiation exists for the intentionality judgments, $F(154) = 3.7, p < .06, \eta^2 = 2\%$.

2. Presentation Format

One of the parameters we explored was presentation format. In many domains of human cognition, making an absolute judgment of a stimulus in isolation is significantly more difficult than making a relative judgment of that stimulus relative to another. Thus, we wondered whether people might be more sensitive to the luck vs. skill difference for moral action if they saw the two conditions (luck and skill) side by side. However, such a presentation format did not alter the results, either for blame ratings or intentionality judgments. If anything, the slight differentiation of intentionality judgments we saw between the two conditions for the aunt's murder in the separate presentations (85% vs. 97%, Table 2) is lost, as the rates are now 95% and 97%, $F < 1$. An additional question asked participants what the main difference was between the two scenarios. The free-response answers reveal that about 80% did recognize the presence vs. absence of skill as the critical difference, but apparently that difference was not sufficient to make a difference in their blame and intentionality judgments.

3. Effects of Verb Choice

In the original scenario by Knobe (2003b), the verb chosen to describe the immoral act was *kill*. It may be argued that this term has strong connotations of intentionality and may therefore be biasing people's responses toward intentionality. To examine this possibility we created three verb conditions for the immoral act: Did he "hit his aunt's heart?" ($N = 79$), "shoot his aunt?" ($N = 38$), and "kill his aunt?" ($N = 40$)?

Table 3. Blame ratings for different verbs describing the stimulus action

	<i>Luck</i>	<i>Skill</i>
How much blame does Jake deserve for hitting his aunt's heart ?	4.3	5.5
How much blame does Jake deserve for shooting his aunt ??	5.8	5.8
How much blame does Jake deserve for killing his aunt ?	6.0	5.8

Table 4. Intentionality judgments for different verbs describing the stimulus action

	<i>Luck</i>	<i>Skill</i>
Did Jake hit his aunt's heart intentionally? (Yes, No)	49%	95%
Did Jake shoot his aunt intentionally? (Yes, No)	84%	90%
Did Jake kill his aunt intentionally? (Yes, No)	100%	100%

The results suggest that when the act is formulated as *hitting the aunt's heart* people differentiate luck from skill both in their blaming responses (Table 3), $F(1, 153) = 20.2, p < .001, \eta^2 = 12\%$, and in their intentionality judgments (Table 4), $F(1, 153) = 35.2, p < .001, \eta^2 = 19\%$. In fact, the differentiation that people show between hitting the aunt's heart with skill vs. luck is now almost as large (95% vs. 49%) as the one we saw for hitting the bull's eye (30% vs. 90%).

The question now arises why the particular action verb would matter so much. There are at least three interpretations to consider. First, the words *kill* or *shoot* might have such a strong connotation of intentionality that people's judgments are biased toward ignoring the moderate amount of luck the person's action relied on. However, this explanation seems weakened by the fact that *kill* is more reliably used to indicate an intentional act than is *shoot*.

Second, the verbs *shoot* and *kill* might have such strong connotations of immorality that participants override their cognitive analysis of the problem and react more in light of their moral outrage. This interpretation concurs with Knobe's (2003b) claim that moral sentiments change the folk criteria for intentionality, but it offers a way out of this problem, namely, by choosing precise, behavioral action verbs.

Third, one major difference between the original killing-aunt scenario and the hitting-bull's-eye scenario is that, holding constant the shooting distance, there are many more ways of killing a person (e.g., hitting her head, throat, heart, lungs...) than there are to hit a bull's eye. In other words, the task is less specified in the killing-aunt scenario. Participants may therefore be lenient about the *skill* condition, because Jake's intention may not have been "to hit the aunt's heart" but rather "to shoot her in any way possible." So even if Jake's hand slipped and the shot went wild, he still fulfilled his rather unspecific intention, hence acted intentionally.

One way to directly test this third interpretation is to hold constant the specificity of the intention in each scenario and then test the impact of skill on intentionality judgments. This can be achieved, for example, by featuring Jake shooting a life-size dummy in the contest scenario or by featuring, in a somewhat science-fiction inspired murder scenario, an assassin's intention of shooting the target's brains (e.g., to destroy her memory) but ending up shooting her heart.

5. Effects of Response Options

In the original scenarios, Knobe (2003b) asked participants a dichotomous question: Did Jake perform his action intentionally or not? This dichotomy may veil a more subtle reaction that people have to the scenarios. For example, people may refuse to say that Jake acted unintentionally (it wasn't entirely an accident), so they choose the slightly more fitting, though not ideal option of "he acted intentionally." Of course, this tension exists only in the case of the immoral action and not in the case of the neutral action (the contest scenario). There may still be a difference between the two types of scenario due to morality, but the difference may not need to refer to the concept of intentionality but to people's reluctance to express anything that would indicate they let a murderer get out of his deserved punishment.

As an initial attempt to offer more than two response options, we formulated the following questions ($N = 71$):

Did Jake shoot his aunt **intentionally**?

- No, he didn't shoot her intentionally (he got lucky), but he deserves full blame.
- Not sure whether one can call that intentional, but either way he deserves full blame
- Yes, there is no doubt that he shot her intentionally.

The same options were formulated correspondingly in the contest scenario, and the comparison between the two scenarios yielded the following results for the *Luck* condition (the *Skill* rates are in parentheses):

Table 5. Finer-grained intentionality judgments in the luck condition (with the skill condition in parentheses)

	<i>No</i>	<i>Not sure</i>	<i>Yes, no doubt</i>
Did Jake hit the bull's-eye intentionally? (Yes, No)	71% (6%)	29% (25%)	0% (69%)
Did Jake shoot his aunt intentionally? (Yes, No)	0% (5%)	32% (16%)	68% (69%)

The results show that increased response options did not change the pattern of results, at least not for the formulation using the verb *shoot* (the only one we tested). Even though a third of the participants are “not sure” whether Jake’s shooting was intentional, virtually the same proportion of participants has similar doubts with respect to the intentionality of the contest-winning act. Most important, 68% of people consider Jake’s lucky shooting act intentional when it kills his aunt whereas 0% consider it intentional when it hits the bull’s eye. An important experimental variation would be to present these finer-grained response options in the face of scenarios that vary in verb choice for the stimulus action (e.g., *hitting* vs. *killing*).

6. Interim Conclusions

The findings by Knobe (2003b) were replicated under a number of conditions: in the original formulation and format, when presenting *Luck* and *Skill* conditions side by side, and when three rather than two response options were made available. Under these conditions we see that people judge a neutral act of hitting the bull’s-eye in a contest intentional only if the agent’s shooting skill is evident, whereas they judge an immoral act of shooting one’s aunt as intentional no matter whether the agent’s shooting skill is evident.

However, the precise description under which the action is presented and intentionality judgments are made (i.e., *hitting*, *shooting*, or *killing*) has a substantial effect on both blame and intentionality judgments. In particular, the original finding by Knobe (and the abovementioned replications) all rely on the use of verbs like *shoot* or *kill* and do not replicate when one describes the action at issue in more specific terms such as *hitting the aunt’s heart*. This pattern of results suggests that people’s judgments are sensitive to the precise action under consideration. For shooting or killing a person from a reasonable distance, a certain amount of luck can be ignored as long as the intention was an unspecific “bringing about the aunt’s death.” If, however, one wants to know whether the agent performed a specific (and difficult) act of, say, hitting the aunt’s heart (or hitting the tiny bull’s eye, for that matter), the presence of skill or luck is taken into account. At present, we must therefore be skeptical of the strong thesis that people’s judgments of intentionality are influenced by moral considerations. The findings that show such influence (by Knobe, 2003b, and our replications) may be confounded with an asymmetry of intention-action specificity.

Note

An updated and expanded version of this report is available at <http://darkwing.uoregon.edu/~bfmalle> (Click on Publications).

References

- Knobe, J. (2003a). Intentional action and side effects in ordinary language. *Analysis*, 63, 190-193.
- Knobe, J. (2003b). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*.
- Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, 101-121.