

# Harming is more intentional than helping because it is more probable: A hidden influence of probability on the Knobe effect

Kuninori Nakamura (knaka@seijo.ac.jp)

Faculty of Social Innovation, Seijo University, 6-1-20, Seijo, Setagayaku  
Tokyo 152-0061, Japan

## Abstract

Knobe (2003) demonstrated that people's intentionality judgments in side effects depend on the outcome of the side-effect, indicating that people's judgments of intentionality of action depend on not only the intention of the actor but also on the result of the action. However, on the basis of findings in judgment and decision making (e.g., Harris, Corner, & Hahn, 2009), the current study proposes another hypothesis to Knobe's (2003) results: the participants' intentionality judgments depended on the probabilities of outcomes provided by the action, rather than on the outcomes itself. To test this hypothesis, the present study employed an identical experimental procedure to Knobe (2003), except that it required not only intentionality and probability judgments for outcomes that resulted from the actions of a company president. The results replicated the findings of Knobe (2003) and showed a relationship between probability and intentionality judgment.

**Keywords:** intentionality, Knobe effect, probability, outcome

## Introduction

Imagine a situation in which a chairman of a company considered starting a new program. This program, according to the vice-president of the company, would surely profit the company, but it would also harm the environment. The chairman asserted that he would start the program in order to make a profit and that he did not mind harming the environment. Finally, the program was started, and the environment was harmed. Did the chairman intentionally harm the environment? Consider another situation in which a chairman considered starting a new program that surely would make a profit and help the environment. In this situation, the chairman started the program to make a profit and had no interest in helping the environment, but the program began and the environment was improved. Now do you think the chairman intentionally harmed the environment?

In both situations, the chairman's main intention to start a new program was the same; however, the impression of the chairman's intentionality may be very different. In fact, Knobe (2003) demonstrated that the chairman in the harming situation was considered to be intentionally harming the environment, whereas he was not intentionally helping the environment when he made the more helpful decision. This indicates that ordinary people judge intentionality depending on what the outcome happens to be. This phenomenon is called the Knobe effect (e.g., Nichols & Ulatowski, 2007) or the side-effect effect (Knobe, 2003; Leslie, Knobe, & Cohen, 2006), and has attracted research attention from various fields, including action theory (Mele, 2003; McCann, 2005), social psychology (Malle, 2006),

moral psychology (Hauser, 2006), philosophy of law (Nadelhoffer, 2008), philosophy of language (e.g., Adams & Steadman, 2004), and developmental psychology (Leslie, Knobe, & Cohen, 2006).

According to Pierre (2010), intentionality is defined as "the power of minds to be about, to represent, or to stand for, things, properties, and states of affairs...." This definition clearly shows intentionality as some inner state of mind. It is a drive or cause for action, and its existence is determined before action. However, the Knobe effect clearly shows that the lay understanding of intentionality violates this fundamental definition: people appear to decide intentionality from the value of the outcome. Thus, this effect can be considered as an example that lay people have odd interpretations of intentionality.

Additionally, the Knobe effect contradicts existing models of intentionality judgment (e.g., Forguson, 1989; Heider, 1958; Jones & Davis, 1965; Malle & Knobe, 1997; Shaver, 1985). Originated in the ancient writings of Aristotle (1892/382 B.C.) and then elaborated on by Hume (1978/1740), models of intentionality judgment have assumed inner state of mind as fundamental components for intentionality such as desire or belief. For example, Malle and Knobe's (1997) model that can be positioned as the most comprehensive one specifies five components for intentionality: (1) desire that concerns the goals or purposes of an actor, (2) belief that concerns the actor's thoughts about the consequences of his or her actions, (3) intention, which links desire to action, (4) awareness that represents an actor's state of mind at the time of acting, and (5) skill, which enables the action. These five components consistently represent will or ability inside the actor, and none of them seem to entail effect of outcome on intentionality judgment. Thus, the Knobe effect can be considered a counterexample to the existing models of intentionality judgment.

Several hypotheses that have been proposed to explain the Knobe effect (e.g., Malle, 2006; Mallon, 2008; Nadelhoffer, 2008; Nichols & Ulatowski, 2007) by focusing on factors that affect lay people's understandings of the meaning of intentionality. Some researchers (e.g., Malle, 2006; Nadelhoffer, 2008) have considered the Knobe effect to arise as a bias in the intentionality process that is driven by either moral (Malle, 2006) or emotional (Nadelhoffer, 2008) considerations. Other researchers (e.g., Adams & Steadman, 2004) have focused on the pragmatic aspect of intentionality judgment, in which the perception of intentionality leads to the blaming of the chairman for hurting the environment, and this effect is strengthened in the harm condition. Additionally, Nichols and Ulatowski

(2007) noted individual differences in how interpretations as to whether something is “intentional” contribute to asymmetry in intentionality judgment between the harming and helping conditions.

This study also aims to explore the Knobe effect by pointing out a new factor that would affect intentionality judgment. More specifically, this study proposes that *probability* plays an important role in the Knobe effect. Although this proposition seems to be unexpected in the literature of intentionality judgment, it is naturally derived from the existing studies. The following descriptions explain this proposition more precisely.

Up to now, many studies on judgment and decision making have demonstrated a dependency of probability judgments for future outcomes on their utilities. (for a review, see Harris, Corner, & Hahn, 2009; Krizan & Windschitl, 2007). In terms of normative theory, utilities and probability are separate components for decision making. However, people’s estimate for how probable the outcome would occur is affected by its desirability or subjective value. This idea was suggested by early research on decision-making (Crandall, Solomon, & Kellaway, 1955; Edwards, 1953, 1962; Irwin, 1953; Marks, 1951; Morlock & Hertz, 1964), but controversy has arisen over the necessity to presume interdependence between utility and probability (Edwards, 1962; Fisher & Jungermann, 1996; Weber & Hilton, 1990).

To settle this controversy, Harris et al. (2009) examined the effect of utility on probability under experimentally controlled conditions. In Harris et al.’s (2009) study, participants were shown dot patterns that indicated the alignment of an apple tree in a farm, and they were required to estimate the probability of randomly picking up a poisonous or sour apple from any tree on the farm. The results of this study showed that the participants estimated higher probabilities of picking up poisonous apples than of picking up sour apples, even when they saw the same dot patterns, indicating that the seriousness of the event surely affected their probability judgments. In sum, the effect of utility on probability judgment can be regarded as an established finding. Further, negative events are estimated at higher probabilities than are neutral or positive events (for more discussion, see Fisher & Jungermann, 1996; Krizan & Windschitl, 2007; Weber & Hilton, 1990).

The finding that utility affects probability indicates a link between probability and outcome in the Knobe effect because the manipulations of outcomes may produce a difference in utility between the two conditions. In addition, a link between probability and intentionality has already been suggested by the existing models of intentionality judgment (e.g., Forgas, 1989; Heider, 1958; Jones & Davios, 1965; Malle & Knobe, 1997; Shaver, 1985). As stated above, these models tried to specify certain components of intentionality judgment, and among them, desire and belief have been positioned as the basic components of intentionality by these models (for a review, see Malle & Knobe, 1997). Between these two components,

belief appears to relate to probability because it concerns thinking about the outcome of an action. For example, following Malle and Knobe (1997), belief is defined as “beliefs or thoughts about the consequences of the act or the act itself before it takes place. (pp.106)” This definition clearly relates belief to a prediction of an outcome, reflecting some measure of uncertainty about a future event. Thus, it is natural to interpret the existing models as suggesting an important role of probability on intentionality.

In sum, reviewing the existing literature on decision-making and intentionality judgments suggests a link between probability and intentionality judgments, as well as one between outcome and probability. As a result, there is a possibility that the Knobe effect, in fact, is the result of an effect of probability on intentionality judgment. The purpose of this paper is to test this possibility. To accomplish this, the following three studies were performed. Study 1 adopted almost the same procedure as Knobe (2003) except that it also required participants to give their own subjective probabilities for harming and helping environments. Study 2a replicated the results in Study 1 and quantitatively examined the effect of probability on intentionality judgment via multivariate analysis. Study 2b tried to replicate the results of Study 2a by employing another scenario used in Knobe’s (2003) study. The results of these three studies consistently supported the hypothesis, but they also showed that effect of outcome remains when controlling for the effect of probability on intentionality judgments.

## Study 1

### Design and procedure

Ninety-five undergraduates participated in Study 1 as a way to receive classroom credit. Fifty participants in the harming condition read the following scenario, which was adopted from Knobe (2003) and was translated into Japanese. The scenario read as follows:

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 harm the environment.”

The chairman of the board answered, “I don’t care at all about harming 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 harmed.

Forty-five participants in the helping condition read almost the same scenario, except the word “harm” was changed to “help” in this version of the text. After reading one of these scenarios, participants answered questions concerning the chairman’s intentionality in either harming or helping the environment (“Do you think the chairman intentionally harmed/helped the environment?”) by two forced choice task (“intentional/not intentional”).

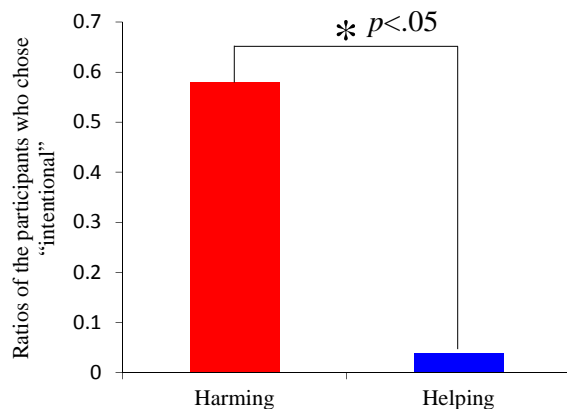


Figure 1. Ratios of participants who chose “intentional” in the harming and helping conditions.

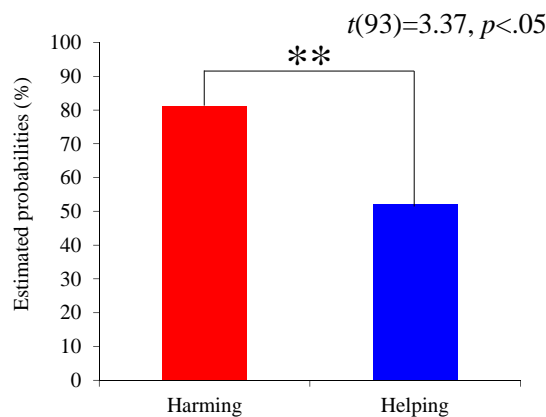


Figure 2. Mean probability estimates for harming/helping the environment.

Additionally, the participants also answered questions concerning the probability for harming/helping the environment that they thought the chairman considered before he made his decision (i.e., “With what probability did the chairman think the program would harm/help the environment?”) by percentage (%). The order of intentionality and probability judgment tasks was counterbalanced.

## Results and discussion

The results shown in Figure 1 clearly demonstrate that the intentionality of the chairman was perceived as being different between the harming and helping conditions. Whereas 60% (27/45) of the participants in the harming condition considered that the chairman intentionally harmed the environment, only 8% (4/50) of the participants in the helping condition considered that the chairman intentionally helped the environment. The difference between the two conditions was statistically significant (ratio test;  $p < .05$ ), indicating that our findings in Study 1 replicated the Knobe effect.

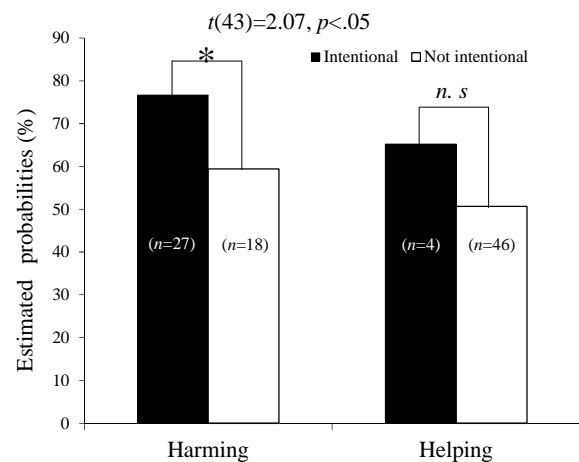


Figure 3. Mean probability estimates as a function of participants’ intentionality judgments.

The results shown in Figure 2 also demonstrate a difference in the probability judgment task between the two conditions. A mean probability estimate in the harming condition was higher than that observed in the helping condition. This difference is statistically significant ( $t(93) = 3.37, p < .05$ ), indicating that the participants considered that the chairman more likely to harm, rather than help the environment.

Figure 3 shows probability estimates as a function of participants’ intentionality judgments, and this appears to indicate that the probability estimates by the participants who chose “intentional” were higher than were those made by the participants who chose “not intentional” as a response. This difference was statistically significant in the harming condition ( $t(43) = 2.07, p < .05$ ). Although the difference was not statistically significant in the helping condition ( $t(48) = 1.20, p > .10$ ), I thought that this was due to the small number of participants who chose “not intentional” in the helping condition ( $n = 4$ ). As far as seeing values of the mean estimates, the results in Figure 3 suggest that intentionality judgment was positively correlated with probability judgment.

The results of Study 1 supported the hypothesis that probability is a crucial factor in the appearance of the Knobe effect. Manipulations toward a certain outcome affected not only intentionality judgment but also probability judgment, and these two judgments correlated with each other. These results suggested that the Knobe effect depends on not only the value of outcome but also on the probabilistic aspect of the scenario.

## Study 2a

Although the results of Study 1 supported the possibility that probability affect intentionality judgment, a quantitative aspect of this relationship still remains uncovered. Thus, Study 2a aimed to explore this point by requiring participants to make intentionality judgment by Likert scale instead of two-forced choice task.

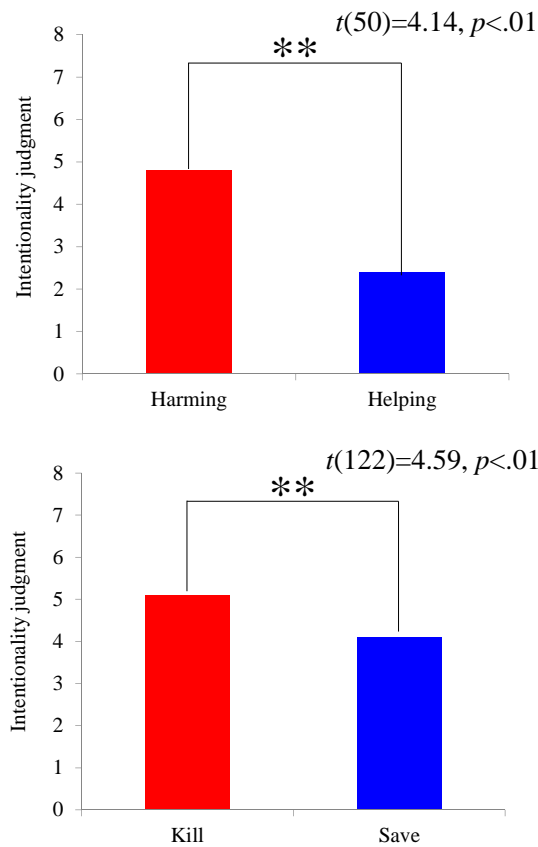


Figure 4. The results of intentionality judgment in Studies 2a (upper graph) and 2b (lower graph).

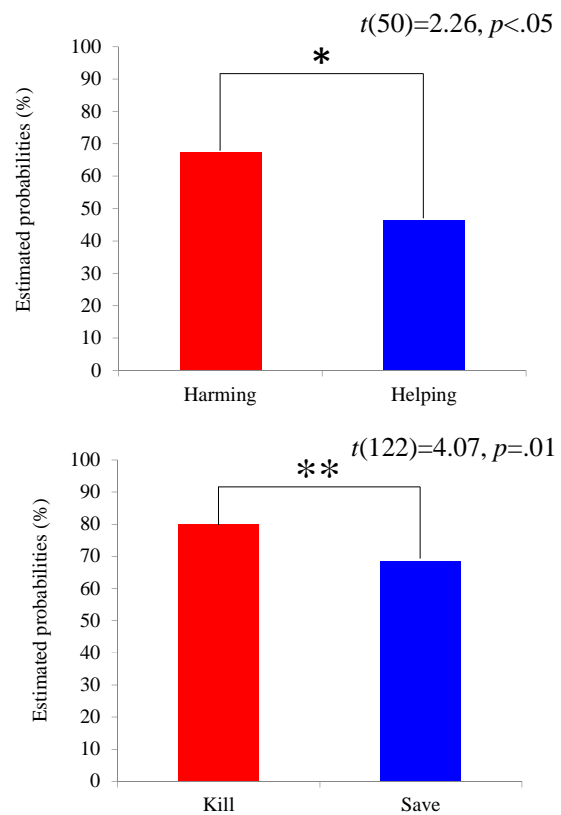


Figure 5. The results of probability judgments in Studies 2a (upper graph) and b (lower graph).

## Design and procedure

Study 2a adopted almost the same procedure as that of Study 1. Twenty-seven participants read the same harming scenario, and twenty-five participants read the same helping scenario that was used in Study 1. Then, after reading one of these two scenarios, the participants completed both the intentionality and probability judgment tasks as described above. However, one difference was that the intentionality judgment in Study 2 was measured on an 8-point scale (0: not intentional; 9: intentional) instead of a two forced-choice task. The order of the two tasks was randomized among all the participants.

## Results and discussion

The results shown in Figures 4 and 5 demonstrate that Study 2 replicated the main results of Study 1. The chairman's intentionality was estimated to be higher in the harming condition than in the helping condition ( $t(50) = 4.14, p < .01$ ), and the mean probability estimate for the participants in the harming condition was significantly higher than the helping condition ( $t(50) = 2.26, p < .01$ ).

To explore the relationships among intentionality and probability judgments, multiple regression analysis was

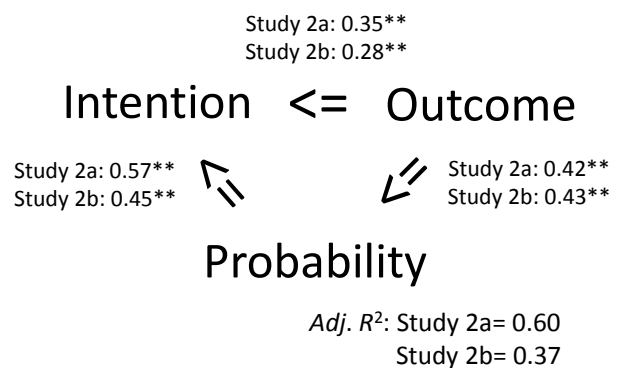


Figure 6. The results of a path analysis conducted in Studies 2a and 2b.

performed. In this analysis, the independent variables were outcome and probability, and the dependent variable was intentionality judgment. The results (Figure 6) showed that both of the independent variables significantly affected intentionality judgment, and the magnitude of the effect from probability was larger than that from value of the result. Additionally, mediation analyses indicate that the indirect effect from outcome via probability on intentionality judgment was marginally significant.

The results of Study 2 indicated the following three things. First, the effect of outcome on probability was robust. Second, the Knobe effect appeared to be a compound phenomenon affected crucially by both probability and outcome. Third, probability appears to be more crucial to the Knobe effect than outcome.

### Study 2b

The results of Studies 1 and 2a are limited in that they used only one aspect of Knobe's research scenario (2003); thus, Study 2b aimed to replicate the findings of both Studies 1 and 2b by using the sergeant scenario that was used in Knobe's (2003) second experiment.

### Design and procedure

Study 2b adopted almost the same procedure as that of Study 2a, except that 124 participants read the sergeant scenario (Knobe, 2003, second experiment), which went as follows:

A lieutenant was talking with a sergeant. The lieutenant gave the order, "Send your squad to the top of Thompson Hill."

The sergeant said, "But if I send my squad to the top of Thompson Hill, we'll be moving the men directly into the enemy's line of fire. Some of them will surely be killed!"

The lieutenant answered, "Look, I know that they'll be in the line of fire, and I know that some of them will be killed. But I don't care at all about what happens to our soldiers. All I care about is taking control of Thompson Hill."

The squad was sent to the top of Thompson Hill. As expected, the soldiers were moved into the enemy's line of fire, and some of them were killed.

The above scenario was used for the "kill" scenario condition in which 62 individuals participated. The remaining 62 participants read the "save" scenario, where word "kill" was exchanged with "save." After reading one of the two scenarios, the participants answered questions concerning intentionality ("Do you think the lieutenant intentionally killed/saved the squad?"), and they completed the probability judgment task ("With what probability did the lieutenant think the squad would be killed/saved?"). These were the same types of questions, asked in the same way, as in Study 1a. The order of probability and intentionality judgment tasks was randomized among the participants. All the participants completed the tasks within 10 minutes.

### Results and discussion

The results shown in Figures 4 and 5 demonstrated that Study 2b replicated the findings of Study 2a. The

intentionality attributed to the lieutenant was estimated to be higher in the harming condition than in the helping condition ( $t(122) = 4.59, p < .01$ ), and the mean probability estimate that the lieutenant thought about killing his soldiers first in order to gain his ultimate win was significantly higher than the mean estimate observed in the helping/saving condition ( $t(122) = 4.07, p < .01$ ). Additionally, the results of a path analysis showed almost the same trends as in Study 2a, which suggested that not only the outcome but also the probability affected intentionality judgment, and the effect of probability is larger than the effect of outcome. These results indicated that the findings in Study 2a are not limited to the chairman scenario.

### General discussion

The results of the above three studies consistently supported the hypothesis that probability plays an important role in the Knobe effect. Study 1 showed significant differences in probability judgments between the harming and helping conditions and between participants who chose an "intentional" responses compared to those who did not. Study 2a found that both the effects of outcome and probability on intentionality judgments were significant, and the effect of probability was larger than that of outcome. Study 2b replicated the results of Study 2a by using another scenario in Knobe's (2003) study. All of these results indicate the crucial influence of probability on the Knobe effect, and are, as a whole, in line with the findings that negative events are estimated to occur with higher probabilities than are neutral or positive events (Harris et al, 2009; Weber & Hilton, 1990).

Previous studies (e.g., Knobe, 2003; Malle, 2006; Mallon, 2008; Nadelhoffer, 2008; Nichols & Ulatowski, 2007) have positioned the Knobe effect as an issue of interpretation in regards to intentionality. In contrast to these studies, however, this paper gives a unique explanation of the Knobe effect. People may consider an action to be intentional because it would make a negative outcome a certainty. The role of probability in the Knobe effect has not been focused on in previous studies, so this explanation uncovers a new aspect of making judgments under uncertain conditions in a Knobe effect situation. Of course, this paper is limited in that it adopts a correlational approach; thus, a causal relation between intentionality and probability remains an unresolved question. Although this paper itself cannot address this question, the author believes that an investigation into the possible close relationship between probability and intentionality is worth further study.

This paper can be also positioned as an example of the influence of belief on intentionality judgment; thus, it supports the existing models for intentionality judgments (e.g., Ferguson, 1989; Heider, 1958; Jones & Davios, 1965; Malle & Knobe, 1997; Shaver, 1985). Of all these previous studies, this current paper corresponds most to Malle and Knobe (1997) because it demonstrates that the Knobe effect can be explained by their model. According to their model,

intentionality judgments are affected by the five factors: desire, belief, intention, awareness, and skill. Knobe (2003) himself considered his effect to be a counterexample of Malle and Knobe's (1997) model because none of these five factors can explain the effect of outcome on intentionality judgments. In contrast to Knobe's interpretation, however, this paper suggests the possibility that the Knobe effect is compatible with Malle and Knobe's model in that it clarifies manipulations toward outcomes according to a specific probability, and this directly corresponds to the effect of the probability itself, as well as to the effect of the belief. This indicates that the Knobe effect is, in fact, supportive of Malle and Knobe's (1997) findings, and it shows that the existing models certainly may explain the Knobe effect by extending a meaning of belief.

This paper is also implicative in revealing that the Knobe effect is a compound phenomenon due to certain outcomes and their probabilities. Knobe (2003) treated his effect as an outcome of intentionality, and subsequent studies (e.g., Mallon, 2008; Nichols & Ulatowski, 2007) assume that outcome is the only factor that can produce the Knobe effect. To this end, they have focused on how to interpret the effect of the outcome. This paper, however, succeeds in pointing out probability as another key factor in the Knobe effect. In addition, when taking into consideration the values of the path coefficients, the findings of this study indicate that probability is a much more important factor than outcome. Of course, this paper does not insist that probability alone can explain the Knobe effect because the effect of outcome was still significant when controlling for the effect of probability. Rather, in line with the existing model for intentionality judgment (e.g., Forguson, 1989; Heider, 1958; Jones & Davios, 1965; Malle & Knobe, 1997; Shaver, 1985), this paper suggests that intentionality judgment should be considered a compound phenomenon, affected by various factors, including desire or belief. Future research should investigate the Knobe effect as result of these and other complex psychological processes.

## References

- Adams, F., & Steadman, A. (2004). Intentional action in ordinary language: core concept or pragmatic understanding? *Analysis*, 64, 173–181.
- Aristotle (1892). *The Nicomachean ethics* (transl. by J. E. C. Welldon). London: Macmillan. (Originally published around 330 B.C.).
- Crandall, V. J., Solomon, D., & Kellaway, R. (1955). Expectancy statements and decision times as functions of objective probabilities and reinforcement values. *Journal of Personality*, 24, 192–203.
- Edwards, W. (1953). Probability preferences in gambling. *American Journal of Psychology*, 66, 349–364.
- Edwards, W. (1962). Utility, subjective probability, their interaction, and variance preferences. *Journal of Conflict Resolution*, 6, 42–51.
- Fischer, K., & Jungermann, H. (1996). Rarely occurring headaches and rarely occurring blindness: Is rarely = rarely? Meaning of verbal frequentistic labels in specific medical contexts. *Journal of Behavioral Decision Making*, 9, 153–172.
- Forguson, L. (1989). *Common sense*. London: Routledge.
- Harris, A. J. L., Corner, A., & Hahn, U. (2009). Estimating the probability of negative events. *Cognition*, 110, 51–64.
- Hauser, M. D. (2006). *Moral minds*. New York: Springer.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Hume, D. (1978). *A treatise of human nature*. L. A. Selby-Bigge & P. H. Nidditch (Eds.). New York: Oxford Univ. Press. (Originally published in 1740.)
- 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. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63, 104–109.
- Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin*, 133, 95–121.
- Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect: Theory of mind and moral judgment. *Psychological Science*, 17, 421–427.
- Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, 101–121.
- Malle, B. F. (2006). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. *Psychological Bulletin*, 132, 895–919.
- Mallon, R. (2008). Knobe versus Machery: Testing the trade-off hypothesis. *Mind & Language*, 23, 247–255.
- Mele, A. R. (2003). Intentional action: Controversies, data, and core hypotheses. *Philosophical Psychology*, 16, 325–340.
- McCann, H. J. (2005). Intentional action and intending: Recent empirical studies. *Philosophical Psychology*, 18, 737–748.
- Nadelhoffer, (2008). Bad acts, blameworthy agents, and intentional actions: Some problems for juror impartiality. *Philosophical Explorations*, 9, 203–219.
- Nichols, S., & Ulatowski, A. J. (2007). Intuitions and individual differences: The Knobe effect revisited. *Mind & Language*, 22, 346–365.
- Shaver, K. G. (1985). *The attribution of blame*. New York: Springer-Verlag.
- Weber, E. U., & Hilton, D. J. (1990). Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 781–789.