

## **Surprising Connections Between Knowledge and Action: The Robustness of the Epistemic Side-Effect Effect\***

James R. Beebe and Mark Jensen (University at Buffalo)

Forthcoming in *Philosophical Psychology*

A number of researchers have begun to demonstrate that the widely discussed ‘Knobe effect’ (wherein participants are more likely to think that actions with bad side-effects are brought about intentionally than actions with good or neutral side-effects) can be found in theory of mind judgments that do not involve the concept of intentional action. In this article we report experimental results that show that attributions of knowledge can be influenced by the kinds of (non-epistemic) concerns that drive the Knobe effect. Our findings suggest there is good reason to think that the epistemic version of the Knobe effect is a theoretically significant and robust effect and that the goodness or badness of side-effects can often have greater influence on participant knowledge attributions than explicit information about objective probabilities. Thus, our work sheds light on important ways in which participant assessments of actions can affect the epistemic assessments participants make of agents’ beliefs.

keywords: Knobe effect, experimental epistemology, knowledge attribution, folk psychology

---

\* We would like to thank audiences at Carleton University, East China Normal University, Grand Valley State University, Kyoto Sangyo University, Kyoto University, Peking University, University at Buffalo, University of Edinburgh, University of South Alabama, and University of Tokyo for feedback we received on previous versions of this paper. We would also like to thank the following members of the Experimental Epistemology Research Group at the University at Buffalo for the contributions they made to collecting the data reported in this article: Joanna Battaglia, Candace Filipski, Joshua Helak, Isaac Kramer, Tyler Manley, Brian McCulley, Robert Schifferli, Brendan Shannon, Matthew Sherry, and Jarrett Steffen.

## 1. Introduction

One of the earliest and most widely discussed results in experimental philosophy is the ‘Knobe effect’ or ‘side-effect effect,’ in which Joshua Knobe (2003a, 2003b, 2004b) found that participants are significantly more likely to say that a bad side-effect was brought about intentionally than a good one. The same kind of result has been found when participants are asked whether a protagonist *desired* that the side-effect be brought about (Tannenbaum, Ditto & Pizarro, 2007), *decided* to bring it about, *was in favor of* or *opposed to* it, or *advocated* the result (Pettit & Knobe, 2009). James Beebe and Wesley Buckwalter (2010) extended this research into the epistemological domain using Knobe’s original vignettes:

The vice-president of a company went to the chairman of the board and said, “We are thinking of starting a new program. We are sure that it will help us increase profits, and it will also *help/harm* the environment.” The chairman of the board answered, “I don’t care at all about *helping/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 *helped/harmed*.

In a between-subjects design, participants were given either the help or the harm version of the abovementioned vignette and were asked “Did the chairman know that the new program would *help/harm* the environment?” Participants indicated their responses using a seven-point Likert scale, ranging from -3 (labeled as ‘the chairman didn’t know’) to 3 (labeled as ‘the chairman knew’). Their results are represented in Figure 1.

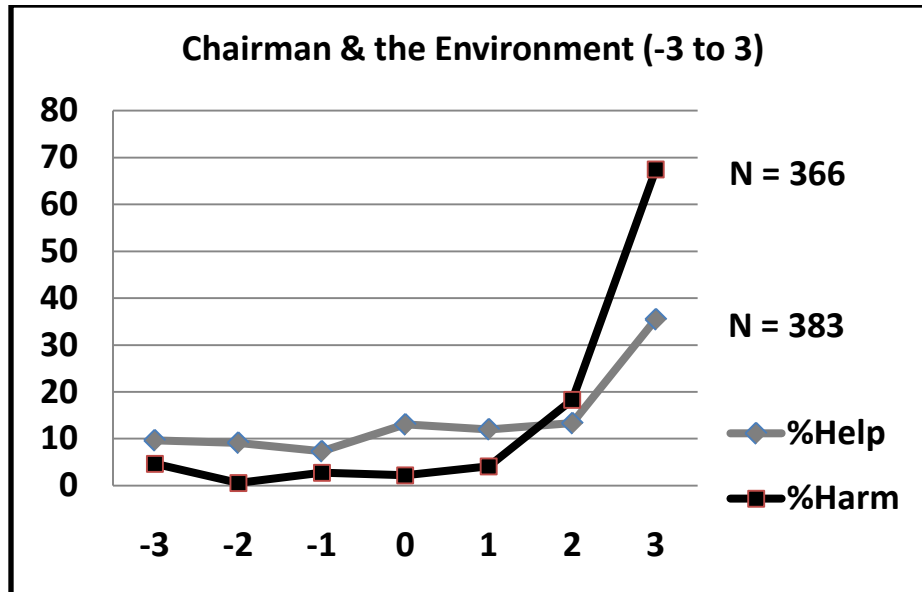


Figure 1

Almost twice as many participants chose the strongest possible affirmation of the chairman's knowledge (viz., response '3') in the harm case (67.5%) as in the help case (35.5%). The number of participants who chose responses 1, 2 or 3 (329) in the harm case was significantly greater than the number of participants who chose 1, 2 or 3 in the help case (233). The distributions about the two mean responses (2.25 and 0.91) were also significantly different.

Various questions arise from the Beebe and Buckwalter study, such as:

- Is the effect large enough to be of any theoretical significance?
- Can the epistemic side-effect effect (as they call their result) be found in as wide a variety of cases as the original side-effect effect?
- To what extent are negative opinions about corporative executives responsible for generating the epistemic side-effect effect?

- Since it seems easier to harm the environment than to help it, are participant responses being guided by this apparent difference in probability or are they being influenced more by the goodness or badness of the side-effect itself?

In the present article we report the results of five experiments designed to address these questions.

Section 2 describes the results of an experiment that deals with a certain kind of concern some have expressed about the theoretical significance of the epistemic side-effect effect. Section 3 reports results from a second experiment that uses vignettes from previously reported studies of the original side-effect effect in an effort to investigate the robustness of the epistemic side-effect effect across a variety of cases. Section 4 describes our attempt to examine the influence that good or bad character trait descriptions might have on epistemic assessments. Sections 5 and 6 report the results of two experiments that investigated participants' assessments of the probabilities of side-effects and the ways these assessments might interact with participants' perception of the goodness or badness of those side-effects. We show that there is good reason to think that the epistemic side-effect effect is a theoretically significant and robust finding and that the goodness or badness of side-effects appears to have more of an influence on participant knowledge attributions than explicit information about objective probabilities.

After demonstrating that the Knobe effect can be found using a variety of folk psychological concepts, Pettit and Knobe (2009) conclude:

It seems to us that there is now good reason to believe there are no concepts anywhere in folk psychology that enable one to describe an agent's attitudes in a way that is entirely independent of moral considerations. The impact of moral judgments, we suspect, is utterly pervasive. (p. 603)

We contend that our results support the hypothesis that the factors responsible for generating the Knobe effect have an utterly pervasive effect on folk psychological attributions, although we do not believe these factors are limited only to moral considerations.

## 2. Experiment 1

One question raised by some who were initially skeptical about the significance of Beebe and Buckwalter's finding concerns the distribution of the mean participant responses across the seven-point scale they used. Mean responses are the most commonly reported type of data in experimental philosophy, despite the fact that other ways of organizing and reporting data are almost always much more useful. In the Beebe and Buckwalter study, the mean response in the help condition was 0.91, while the mean in the harm condition was 2.25 (cf. Figure 2).

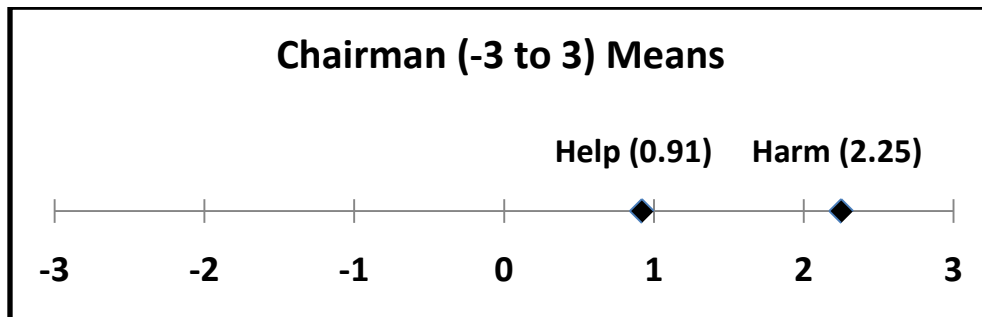


Figure 2

A few skeptics suggested that because the means do not straddle the mid-point of the Likert scale, the results were less theoretically significant than they might have otherwise been. To address this concern, we conducted a follow-up study that replicates Beebe and Buckwalter's findings in important ways.

## 2.1. Methods

### 2.1.1. Participants

185 undergraduate college students from a large, public university in the northeastern United States participated in this study on a voluntary basis. They were neither paid nor given course credit for their participation.

### 2.1.2. Materials and procedure

In a between-subjects design, 87 participants were given one of the two vignettes used in the Beebe and Buckwalter study but were asked to indicate their responses in a forced-choice format, selecting either ‘The chairman knew’ or ‘The chairman didn’t know.’ Another set of 98 participants were given the same vignettes but were directed to register their responses on a Likert scale that ranged from 1 to 7 instead of from -3 to 3.

## 2.2. Results

In response to the forced-choice question, 68% of participants who received the harm version of the vignette indicated that they thought the chairman knew the environment would be harmed, whereas only 16% of those who received the help version thought that the chairman knew it would be helped (cf. Figure 3).

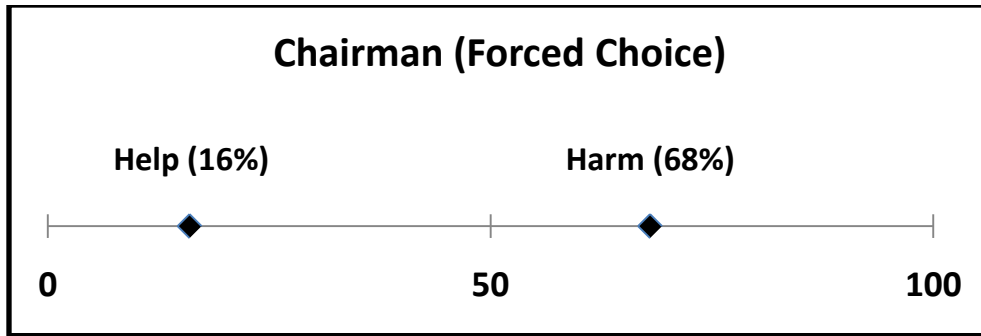


Figure 3

A chi-square test of goodness-of-fit was performed to determine whether subjects responded equally in the forced choice scenario. Responses were not equally distributed,  $X^2(1, N=87) = 22.96, p < 0.001$ . Among participants who received the 1 to 7 response scale, the mean response of those in the help condition was 3.35, while the mean response of those in the harm condition was 6.37 (cf. Figure 4).

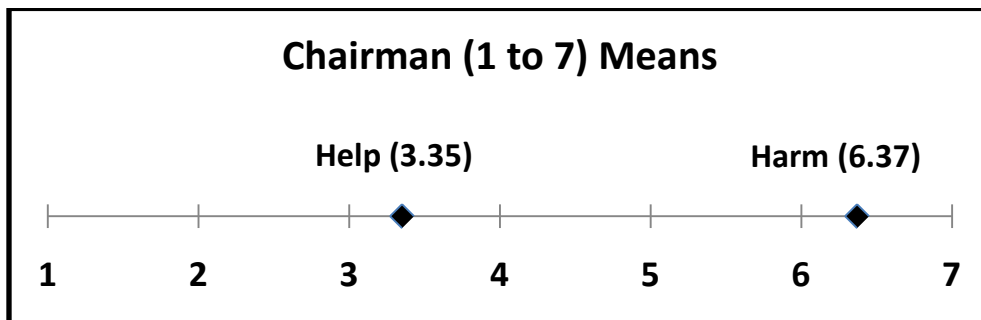


Figure 4

The data points in Figures 3 and 4 clearly straddle their respective midpoints. The full histogram for the data from the 1 to 7 cases appears in Figure 5:

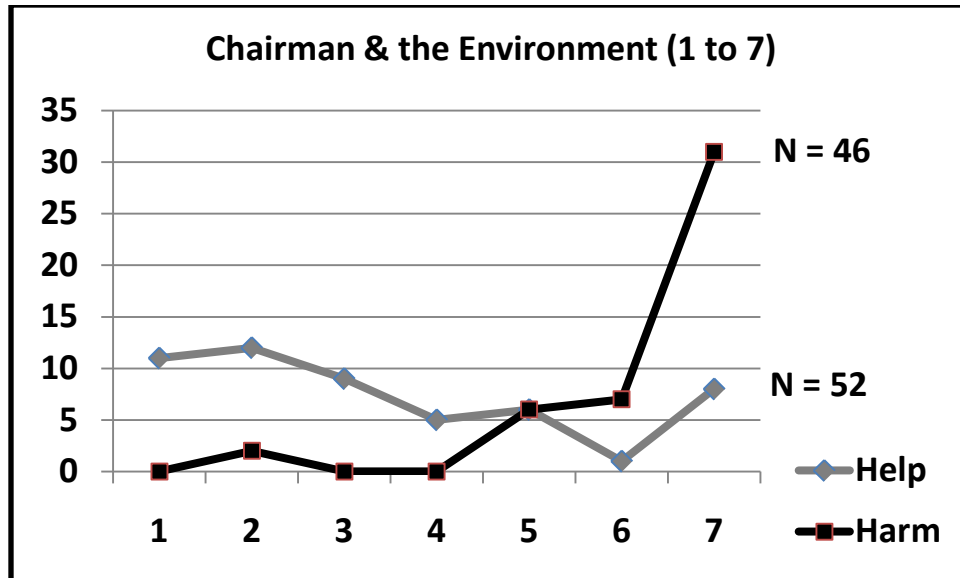


Figure 5

A two-tailed (parametric) t-test ( $t(83) = -9.047, p < 0.0001$ ) and a (non-parametric) Mann-Whitney U test ( $U(98) = 2066.5, z = 6.20, p < 0.001$ ) revealed that the difference in responses was significant. When we compare the data represented in Figures 1 and 5, we find that the data series from the two harm conditions are roughly equivalent ( $t(65) = -0.605, p > 0.05$ ) but that the data series in the two help conditions differ significantly from one another ( $t(66) = -5.123, p < 0.00001$ ). In addition, a significantly greater percentage of participants in the help condition chose to use the left-hand side of the Likert scale when the scale ran from 1 to 7 than when it ran from -3 to 3 ( $t(59) = 5.2, p < 0.00001$ ).<sup>1</sup>

### 2.3. Discussion

The central finding from Experiment 1 is that the key data points from each part of the experiment clearly straddle the mid-points of their respective response continua. In fact, in the 1

<sup>1</sup> Participant responses on the left-hand side of the midpoints in both studies were given new scores of 0, and responses on the right-hand sides were given new scores of 1. Midpoint responses were ignored, and the distributions of the remaining data were compared.



to 7 study—which is formally equivalent to Beebe and Buckwalter’s original study—the spread between the data points is significantly greater than in the original. In spite of this difference, it seems that the two studies should be getting at exactly the same underlying phenomena.

Other researchers have found similar differences between the responses given to Likert scales that include only positive numbers and scales that include both positive and negative numbers. Norbert Schwartz (1995), for example, asked two sets of participants “How successful would you say you have been in life?” and directed them to respond using one of two eleven-point scales. The extreme ends of both scales were labeled as ‘Not at all successful’ and ‘Extremely successful,’ but one scale ran from 0 to 10, while the other ran from -5 to 5. Schwarz found that 34% of the respondents chose a value below the mid-point of the 0 to 10 scale, but only 13% chose one on the -5 to 5 scale. Schwartz (1995) hypothesizes:

When [‘not at all successful’] is combined with the numeric value ‘0,’ respondents interpret it to refer to the absence of noteworthy success. However, when the same label is combined with the numeric value ‘-5,’ they interpret it to refer to the presence of explicit failure. (p. 157)

Something similar may be responsible for the difference between the distribution of data points in Figures 2 and 4. The value ‘-3’ may suggest the presence of explicit epistemic failure, while ‘1’ in the 1 to 7 study may suggest only a low degree of epistemic success. Because the chairman’s belief in the help condition does not have features that make it particularly poor from an epistemic perspective, participants may be reluctant to choose ratings from the left-hand side of a scale that ranges from -3 to 3. But they may feel more freedom to use the left-hand side of a scale that does not imply the presence of negative features.<sup>2</sup>

---

<sup>2</sup> Cf. Cullen (2010) for further discussion of this issue.

One lesson from the above exercise is that, while asking whether mean values lie on either side of a certain mid-point may be of some significance, this test is a rather blunt instrument. Whether data points straddle a mid-point may simply be the result of how participant responses are restricted or of unintended pragmatic features of questionnaire formats that have no genuine significance. These results caution against setting up simplistic tests for theoretical significance when only sample means are used. Of course, one way of avoiding many of these problems is simply to place less stock in means as useful and significant ways of reporting experimental data. Compare the volume of information communicated by Figures 1 and 5—which do not employ means—to that communicated by Figures 2 and 4—which employ only means.<sup>3</sup>

### **3. Experiment 2**

To see whether the epistemic side-effect effect can be found in a wide variety of cases, like the original side-effect effect, we performed a second experiment that employed research materials from previously published studies of the latter effect. The epistemic side-effect effect was found in each case.

#### 3.1. Methods

##### 3.1.1. Participants

619 undergraduate college students from a large, public university in the northeastern United States participated in this study on a voluntary basis. They were neither paid nor given course credit for their participation.

---

<sup>3</sup> For further discussion of the issue of whether data points straddle a midpoint, cf. Pettit and Knobe (2009, 589-590).

### 3.1.2. Materials and procedure

In Knobe's (2003a, 2003b, 2004b) original investigations into the side-effect effect, emphasis was placed primarily on the role that moral considerations play in shaping attributions of intentionality. In an effort to demonstrate that more general evaluative considerations can have similar effects, Knobe (2004a) constructed a pair of cases in which the goodness or badness of the side-effect of a protagonist's action was aesthetic rather than moral. In a between-subjects design, we gave 234 participants either the *better* or *worse* version of the following, modified version of Knobe's (2004a) vignette:

The Vice-President of a movie studio was talking with the CEO. The Vice-President said: "We are thinking of implementing a new policy. If we implement the policy, it will increase profits for our corporation, but it will also make our movies *better/worse* from an artistic standpoint." The CEO said: "Look, I don't care one bit about making our movies *better/worse* from an artistic standpoint. All I care about is making as much profit as I can. Let's implement the new policy." They implemented the policy. Sure enough, the policy made the movies *better/worse* from an artistic standpoint.

In Knobe's original version of the vignette, the CEO's reply to his subordinate includes the following self-attribution of knowledge: "I know that we'll be making the movies worse from an artistic standpoint." In order to test for the effect of evaluative considerations on attributions of knowledge, we omitted this avowal. Participants were asked to indicate on a seven-point Likert scale the extent to which they agreed or disagreed with the following claim: "The CEO knew that the new policy would make the movies *better/worse* from an artistic standpoint." '-3' was labeled 'Strongly Disagree,' '0' 'Neutral' and '3' 'Strongly Agree.'

In the case Knobe first used to demonstrate the side-effect effect, the chairman seems to deserve no praise when his decision to implement a new program helps the environment, but he does seem to deserve blame when his decision results in a harm to the environment. Thomas Nadelhoffer (2004a) hypothesizes that the original side-effect effect findings can be explained primarily as the result of the distorting influence of feelings of blame in the harm case. In an effort to show that feelings of blame are not the primary factor driving the effect, Knobe and Mendlow (2004, p. 24) constructed a vignette in which the protagonist brings about a bad or harmful side-effect but is not blameworthy for doing so.<sup>4</sup> They found that the badness of an action can still generate high attributions of intentionality in such a case. We presented 139 participants with either the *increase* or the *decrease* version of the following, modified version of Knobe and Mendlow's (2004) vignette:

Susan is the president of a major computer corporation. One day, her assistant comes to her and says, "We are thinking of implementing a new corporate restructuring plan. It will simplify our corporate structure, *and/but* it will also *increase/decrease* sales in New Jersey for the next quarter." Susan says "I don't care about what happens in the next quarter. We need to simplify our corporate structure. Let's implement the new plan." They implemented the new plan. Sure enough, sales in New Jersey *increased/decreased* in the following quarter.

Participants were asked to indicate on a seven-point Likert scale the extent to which they agreed or disagreed with the claim "Susan knew that the new plan would *increase/decrease* sales in

---

<sup>4</sup> Nadelhoffer (2004b, 259-260) writes, "In their reply to my work, Knobe and Mendlow suggest that I am committed to the position that people only regard side effects as intentional if the agents who bring them about are blameworthy—i.e. on their reading of my view, blame is a *necessary* condition of folk ascriptions of intentional action in side effect cases.... Of course, I do frequently suggest that blame attribution helps explain the biasing effect that moral considerations have on lay persons' judgments about intentional action, but just because I claim that blame helps explain this biasing effect, it does not follow that I must believe that blame either *fully* explains this effect or explains it in all instances."

New Jersey in the next quarter.” Again, ‘-3’ was labeled ‘Strongly Disagree,’ ‘0’ ‘Neutral’ and ‘3’ ‘Strongly Agree.’

In order to provide an initial test of the hypothesis that nonconscious moral judgments can have important effects on attributions of intentionality, Knobe (2007) constructed a pair of cases that were designed to elicit a conflict between participants’ immediate, intuitive judgments and their more reflective, conscious judgments. Knobe claimed to find that unconscious assessments of the goodness or badness of an action are able to generate the pattern of responses distinctive of the side-effect effect. In our study, 246 participants were given either the *fulfill* or the *violate* version of the following, modified vignette from Knobe (2007):

In Nazi Germany, there was a law called the “racial identification law.” The purpose of the law was to help identify people of certain races so that they could be rounded up and sent to concentration camps. Shortly after this law was passed, the CEO of a small corporation decided to make certain organizational changes. The vice-president of the corporation said: “By making those changes, you’ll definitely be increasing our profits. But you’ll also be *violating/ fulfilling* the requirements of the racial identification law.” The CEO said: “I don’t care one bit about that. All I care about is making as much profit as I can. Let’s make those organizational changes!” As soon as the CEO gave this order, the corporation began making the organizational changes.

As in the aesthetic help and harm case above, the CEO’s reply in Knobe’s original vignettes include an explicit self-attribution of knowledge: “Look, I know that I’ll be *fulfilling/violating* the requirements of the law, but I don’t care one bit about that.” Again, we omitted this attribution and asked participants to indicate the extent to which they agreed or disagreed with

the following claim: “The CEO knew that the organizational changes would *violate/ fulfill* the requirements of the law.” Responses were registered on a 7-point scale ranging from -3 to 3.

### 3.2. Results

When Knobe (2004a) presented the case of the movie studio CEO to participants and asked them “Did the CEO intentionally make the movies *better/worse* from an artistic standpoint?” he found that 54% of participants in the aesthetic harm condition indicated that the CEO acted intentionally, whereas only 18% of participants in the aesthetic help condition thought so. Unlike Knobe’s findings, in which the size of the effect of aesthetic evaluations was substantially smaller than in the environment case, our modified version of the case uncovered a pattern of responses that closely mirrors the pattern found in Beebe and Buckwalter’s study (cf. Figure 6).

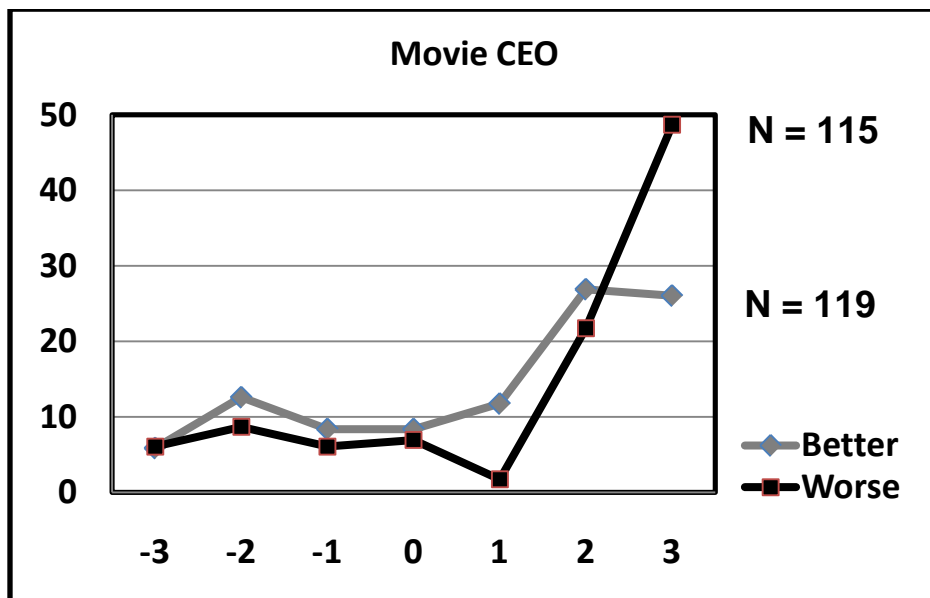


Figure 6

As in the chairman and the environment case, nearly twice as many participants selected the strongest affirmation of the CEO’s knowledge (i.e., response ‘3’) in the ‘worse’ condition

(48.7%) as in the ‘better’ condition (26.1%).<sup>5</sup> Because so many participants selected ‘3’ in the worse condition, the number of participants that selected any of the other answer choices was generally lower in the worse condition than in the better condition. Together these facts produce the pattern of data lines that characterize the epistemic side-effect effect. The mean response in the better condition is 0.92, and the mean response in the worse condition is 1.50. The difference between the distributions about these two means is statistically significant.<sup>6</sup> Unlike Beebe and Buckwalter’s original finding, however, when participant responses of 1, 2 or 3 are given a new score of ‘1,’ responses of -3, -2 or -1 are given a new score of 0, and responses of ‘0’ are ignored, the difference between the number of participants who issued a “positive” affirmation of the CEO’s knowledge and the number of those who gave a “negative” response in the two conditions is not statistically significant ( $t(213) = -1.161, p > 0.05$ ).

In the case of Susan, who either increased or decreased sales in New Jersey for the next quarter, nearly twice as many participants chose response ‘3’ in the harm condition (44.3%) as in the help condition (23.2%) (cf. Figure 7). This difference is statistically significant ( $t(134) = -2.680, p < 0.01$ ).

---

<sup>5</sup> When participant responses of ‘3’ are given a new score of ‘1’ and all other participant responses are given a new score of 0, a two-tailed t-test reveals that the difference between number of participants who selected ‘3’ in the two conditions is indeed statistically significant ( $t(226) = -3.662, p < .001$ ).

<sup>6</sup> T-test result:  $t(231) = -2.204, p < 0.05$ . Mann-Whitney U test:  $U(234) = 8348.5, z = 2.91, p < 0.01$ .

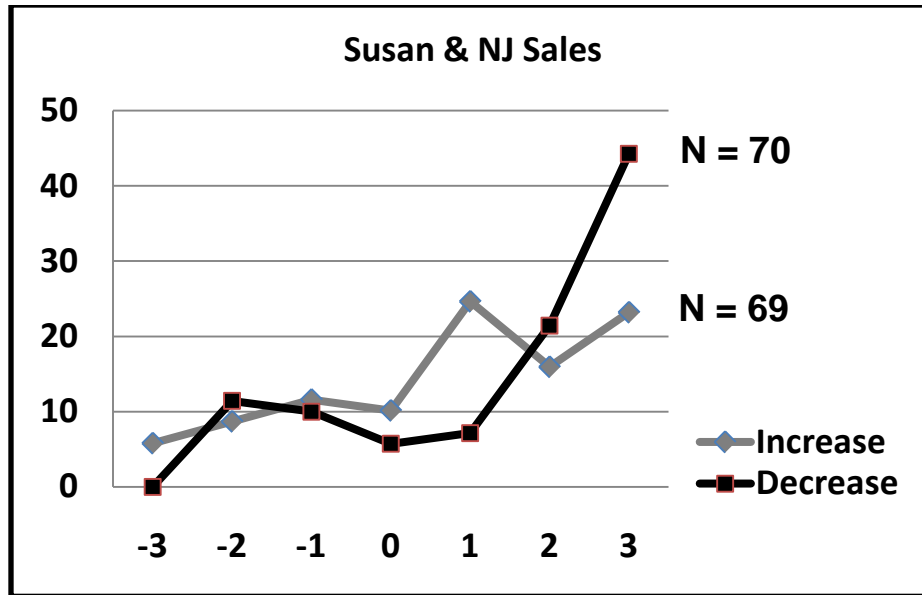


Figure 7

The mean responses were 1.5 (decrease) and 0.80 (increase), and the two distributions were significantly different from one another.<sup>7</sup> As in the movie studio CEO case, the difference between the distributions of positive and negative scores in the two conditions was not significant ( $t(123) = -0.809, p > 0.05$ ).

Although Knobe and Mendlow (2004) intended the case where Susan decreases sales to be one in which a bad side-effect was brought about blamelessly, Mark Phelan and Hagop Sarkissian (2008) performed a replication of Knobe and Mendlow’s study in which they asked participants to rate the degree of praise or blame Susan deserved for decreasing sales and to say whether they thought that decreasing sales was bad. While 94.5% of participants indicated that Susan deserved neither praise nor blame—a finding consistent with Knobe and Mendlow’s prediction—only 14% of participants thought that decreasing sales was bad. Phelan and Sarkissian (2008) conclude, “Thus, whatever explains the high proportion of intentionality judgments in this case, it is presumably not the perceived badness of the side effect.” (p. 295)

<sup>7</sup> T-test:  $t(137) = -2.274, p < 0.05$ . Mann-Whitney U:  $U(139) = 3011.0, z = 2.51, p < 0.05$ .



In Knobe's (2007) original study of the CEO in Nazi Germany, he found that 81% of subjects in the 'violate' condition said that the CEO violated the requirements of the racial identification law intentionally, whereas only 30% in the 'fulfill' condition said that he fulfilled the requirements intentionally. Our study reveals a similar pattern of results (cf. Figure 8).

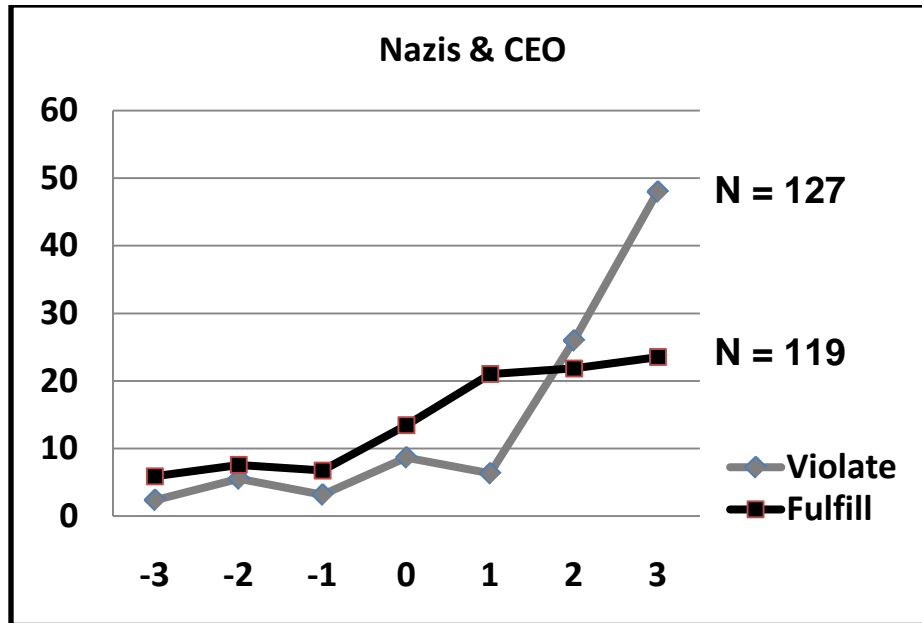


Figure 8

48.0% of participants in the violate condition chose response '3,' whereas only 23.5% in the fulfill condition chose response '3.' This difference is statistically significant ( $t(242) = 4.138, p < 0.001$ ). The mean responses in the two conditions were 1.81 (violate) and 0.96 (fulfill), and the difference between the two distributions is statistically significant.<sup>8</sup> In this case, as in the chairman and the environment case, the difference between the distribution of participant responses on either side of the '0' in each condition was statistically significant ( $t(191) = 2.172, p < 0.05$ ).

<sup>8</sup> T-test:  $t(238) = 3.906, p < 0.001$ . Mann-Whitney U:  $U(246) = 9922.0, z = 4.24, p < 0.001$ .

### 3.3. Discussion

The results of Experiment 2 show that the epistemic side-effect effect is not due to features that are unique to the original chairman and the environment case. When the chairman harms the environment, he violates ethical and social norms. However, when the movie studio CEO makes movies that are artistically poorer or Susan decreases sales in New Jersey, they do not. And yet we find the same basic pattern of responses. The movie studio case shows that aesthetically bad side-effects can generate the epistemic side-effect effect. And while the side-effect of Susan's action in the decrease condition may be financially bad in some sense, because her action is aimed at a greater corporate good, she seems quite justified in simplifying her company's corporate structure. We thus see that the epistemic side-effect effect can be found in cases of blameless, morally neutral harms. The CEO of the corporation in Nazi Germany is also morally blameless for violating an unjust law, even though he will obviously not be blameless in the eyes of local authorities. The CEO in the 'violate' condition still seems to count as bringing about a side-effect that is "bad" in some sense because he has broken a law, which is normally (at least in the type of society whose laws figure most saliently in the minds of our participants) a bad thing to do.

In each of the pairs of cases used in this experiment, a character in a key decision-making role receives testimony from a subordinate about a possible course of action. In both the help and harm (or good and bad) versions of each pair, the testimony that is given seems to be epistemically on a par. If there is reason to trust the subordinate's testimony in one version, there ought to be reason to trust the testimony in the other. And since it seems that epistemic assessments ought to be made solely on the basis of the quality of evidence or reasons available

to a subject, one might expect participants to make symmetric attributions of knowledge in the above pairs of cases. However, participants seem to be strongly influenced by apparently non-epistemic factors in deciding whether or not someone knows.

#### **4. Experiment 3**

One worry that some have expressed in response to the above findings concerns the fact that the chief protagonists in the environment, movie, New Jersey and Nazi cases are all CEOs, and CEOs tend to be viewed rather poorly in modern society. The idea is that (a) people tend to assume that CEOs are unscrupulous and will eagerly cast aside moral and legal norms if they can make an extra buck by doing so and (b) this (alleged) fact may be providing participants in the harm conditions with an extra “push” toward saying that the chief executives knew that the bad outcomes would occur. If these (perhaps largely unconscious) negative assessments of the character of the chief protagonists in the above cases are at least partly responsible for the epistemic side-effect effect, then the goodness or badness of the side-effects themselves may be playing a less significant role than Knobe and many others tend to think. In order to test this suggestion we ran an experiment that employed two pairs of cases in which information about the characters of the chief protagonists was manipulated.

##### 4.1. Methods

###### 4.1.1. Participants

390 undergraduate college students from a large, public university in the northeastern United States participated in this study on a voluntary basis. They were neither paid nor given course credit for their participation.

#### 4.1.2. Materials and procedure

Steven Guglielmo and Bertram Malle (in press, n. 4) ran a study of the side-effect effect in which they describe the character of a city official in a soft way in one condition (e.g., “known for her kind heart and her ability to serve the best interests of all of the city’s residents”) and a harsh way in another (e.g., “known for her harsh governing style”). The vignettes of both conditions went on to describe the mayor undertaking the same course of action that had the same harmful side-effect. Participants were asked whether the mayor had intentionally brought about the harm in question. Following Guglielmo and Malle, we presented 223 participants with either the negative character description or the positive character description below, followed by a bit of common information:

**Negative character condition:** Mayor Emily Spires had always been known for her harsh governing style. Spires had a reputation as a nasty politician with a low tolerance for dissent and little concern for the city’s lower-class folks. In past years, she had taken some heat for not being supportive enough of social programs that would benefit her less wealthy constituents, instead focusing on policies backed by those who financially backed her. Recently, Spires received an offer from a local company to build a franchise in her city. This was, by all accounts, an offer that she could not refuse. She knew that having the company in the city would attract many other businesses in the near future, and would provide a much-needed boost for the city.

**Positive character condition:** Mayor Emily Spires had always been known for her kind heart and her ability to serve the best interests of all of the city’s residents. She was well-liked by people of all political affiliations and she did her best to try to deal with the

problems and concerns of all her constituents, both upper- and lower-class. In past years, she was often recognized by the local newspaper for her effective and unifying leadership style. Recently, Spires received an offer from a local company to build a franchise in her city. This was, by all accounts, an offer that she could not refuse. She knew that having the company in the city would attract many other businesses in the near future, and would provide a much-needed boost for the city.

**Common information:** However, supporting this project—and devoting money toward it—would mean that funding for various city services, including a program that provided meals for the homeless, would be reduced. It was a tough call, but Spires couldn't be concerned with funding for the homeless right now. Forging ahead with the plan to attract this company would be best for the city. So, she decided to fund the project. Sure enough, the company ended up moving to the city, but the meal program took a funding cut, and the fate of the homeless was uncertain. After this, most were able to get by, finding food here and there, despite the cuts to the meal program. But one homeless man, Tom Patton, was hit particularly hard by the funding cut. Unable to find food on his own, Patton was found by the police—he had died of starvation.

As in some of the other vignettes described above, Guglielmo and Malle's original vignettes included an explicit attribution of knowledge to the chief protagonist: "Knowing that the homeless might suffer as a result, she decided to fund the project." This was omitted and participants were asked "Did Spires know that the homeless would go hungry as a result of her plan? Circle the number that best represents the situation described above." Participants were directed to indicate their answer on a seven-point Likert scale, where '-3' was labeled 'Spires didn't know' and '3' 'Spires knew.'

Edouard Machery (2008) has offered an explanation of the side-effect effect that is based upon the idea that protagonists construe bad side-effects as costs that must be incurred in order to receive certain benefits. Ron Mallon (forthcoming) has attempted to show that Machery's hypothesis is mistaken by demonstrating that the side-effect effect can be found in the following pair of vignettes in which the protagonist does not view the bad side-effect as a cost:

**Harmful Gang Leader:** A member of a local gang went to the leader and said, "We are thinking of trying a new tactic. It will flood the neighborhood with cheaper cocaine, increasing our profits, but *it will also harm the cops since more cops will die in drug-related violence.*" The leader answered, "I admit that *it would be good to harm the cops,* but I don't really care about that. I just want to make as much profit as I can. Let's implement the new tactic." They did implement the new tactic, and sure enough, *the cops were harmed since more cops died in drug-related violence.*

**Helpful Gang Leader:** A member of a local gang went to the leader and said, "We are thinking of trying a new tactic. It will flood the neighborhood with cheaper cocaine, increasing our profits, but *it will also help hard-core addicts to have more money for food and housing.*" The leader answered, "I admit that *it would be good for people to have more money for food and housing,* but I don't really care about that. I just want to make as much profit as I can. Let's implement the new tactic." They did implement the new tactic, and sure enough, *hard-core addicts were helped by having more money.*

In our study 167 participants were given either the help or the harm version of the gang leader case and were asked "Did the gang leader know that the hard-core addicts would be helped?" or "Did the gang leader know that the cops would be harmed?" In a forced-choice format,

participants were directed to select either “The gang leader knew” or “The gang leader didn’t know” as their answer.

#### 4.2. Results

Guglielmo and Malle found that when Mayor Emily Spires’ character was described in a negative fashion, she was more often judged to have let the homeless go hungry intentionally (51%) than when her character was described positively (26%). In our study, however, there was no significant difference between the distributions about the mean responses in the kind (0.98) and harsh (1.15) conditions ( $t(220) = -0.730, p > 0.05$ ) (cf. Figure 9).<sup>9</sup>

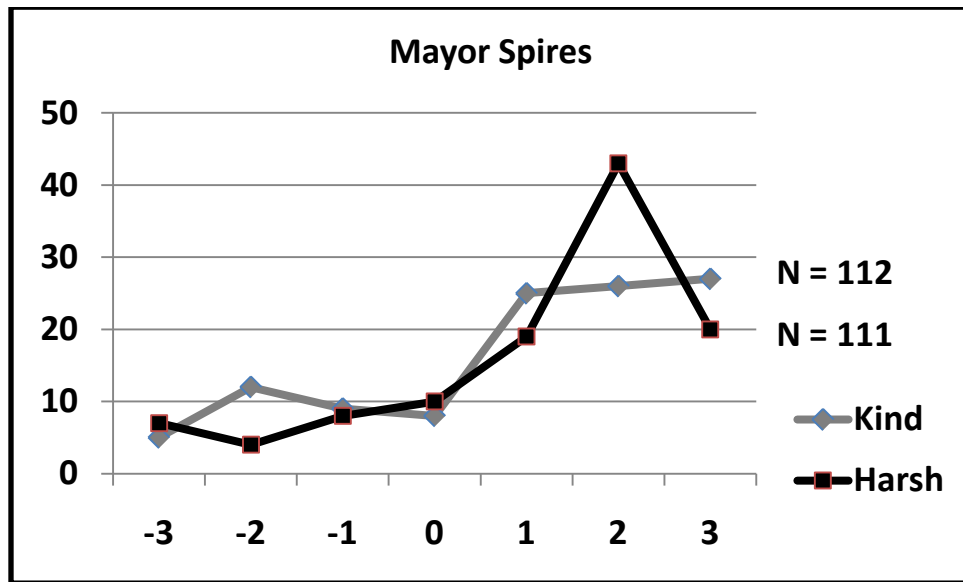


Figure 9

Pairwise t-tests reveal that the responses in both of the Spires conditions are not significantly different from the responses in each of the “good” conditions of the environment, movies, New Jersey, and Nazi cases reported above.

<sup>9</sup> A Mann-Whitney U test also failed to find significance:  $U(223) = 6483.0, z = 0.55, p > 0.05$ .

When Mallon (forthcoming) presented participants with the gang leader case and asked them whether the leader intentionally brought about the side-effect in question, 62% of those who read the story about the harmful gang leader thought that he brought about the harm intentionally, whereas only 28% thought that the helpful gang leader brought about the help intentionally. In a similar fashion, 93% of the participants in our study who read about the harmful gang leader thought that he knew the harm in question would come about, whereas only 59% of those who read about helpful gang leader thought that he knew that the help would come about (cf. Figure 10). According to a chi-square test of goodness-of-fit, the responses were not equally distributed,  $X^2(1, N=167) = 23.72, p < 0.001$ .

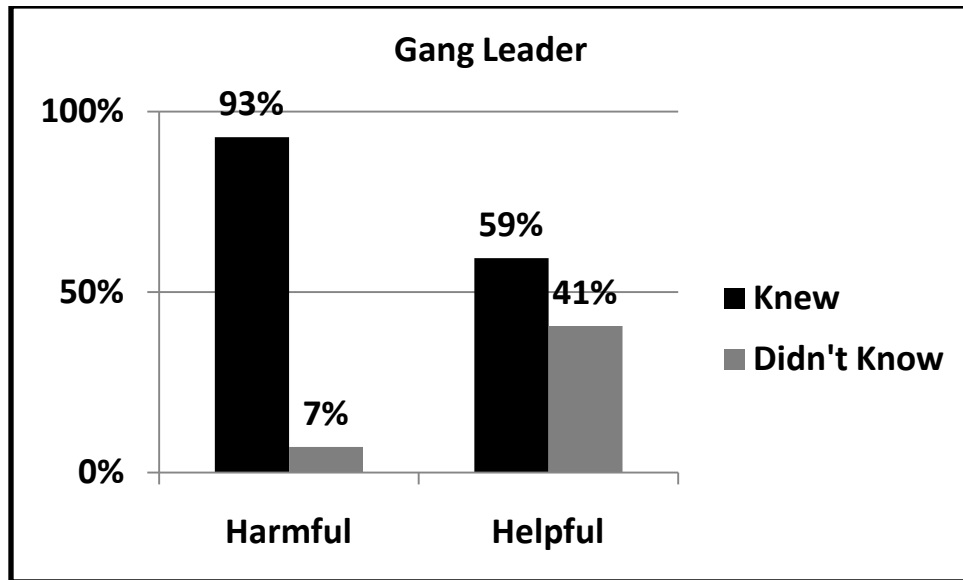


Figure 10

#### 4.3. Discussion

The hypothesis that tacit views on the part of research participants about the unscrupulous characters of CEOs might be responsible for much of the epistemic side-effect effect predicts (a)



that the epistemic side-effect effect should be found in the Mayor Spires case and (b) that high attributions of knowledge should be found in both the help and the harm versions of the gang leader vignette. Neither prediction was confirmed by the present study. (Of course, failing to find confirmation for a hypothesis should not be taken to be the same as succeeding in finding disconfirmation.)

One explanation for the results we found in our Mayor Spires case is that information about the good or bad character of a protagonist is simply not sufficient to generate the epistemic side-effect effect. However, because the Mayor Spires vignettes were longer than any of the others reported in this article, the research materials might have been less able to uncover an effect than other kinds of materials. Further empirical investigation of this possibility is required before any conclusions about the power of character descriptions to influence epistemic assessments can be drawn with any reasonable degree of confidence.

If information about corrupt moral character leads participants to be more willing to attribute knowledge, it seems that we should expect high attributions of knowledge in both versions of the gang leader case. The protagonist is the leader of a drug-dealing gang and in both conditions is taking steps to flood a neighborhood with a harmful narcotic. Contrary to the hypothesis under consideration, however, these two items of information do not generate high attributions of knowledge in both conditions. Although only preliminary conclusions can be drawn from this brief study, when one places the present results against the backdrop of earlier results involving the epistemic side-effect effect, it seems that the goodness or badness of a protagonist's character is unlikely to be as responsible for the epistemic side-effect effect as information about the goodness or badness of the side-effect itself.

On a separate note, the results of our gang leader study provide further confirmation of Mallon's (forthcoming) claim that the protagonist's assessment of the goodness or badness of a side-effect is not what is important for the general class of side-effect effects. The harmful gang leader thinks it would be good to harm the cops in his area, but participants appear to be guided by their own assessments of the merits of this side-effect.

## **5. Experiment 4**

In section 3.3 above, we remarked that the participants in our studies might not be making epistemic assessments solely on the basis of epistemic factors such as evidence or reasons but might instead be influenced by non-epistemic factors. From a reflective epistemological viewpoint, it seems wrong to make the question of whether the movie studio's belief counts as knowledge depend upon the aesthetic properties of the movies he is making. Instead of taking a somewhat negative perspective on the abilities of ordinary participants to make proper epistemic assessments, some scholars have offered the following, more optimistic explanation of the situation.<sup>10</sup>

The explanation begins by noting that it is in general easier to harm the environment than to help it, to make bad (or worse) movies than to make good (or better) ones, and to decrease sales rather than to increase them. But if it is easier to bring about the harmful side-effects in question, this means that the probability of succeeding in bringing about a harm should be greater than the probability of succeeding in bringing about a help. In determining whether the chief protagonist in each vignette knows that a certain side-effect will be brought about, it would be eminently reasonable for participants to take into account the relative probabilities of these side-effects actually occurring. But if participants are being guided by rationally grounded

---

<sup>10</sup> Thanks to Berislav Marusic for bringing the following line of response to our attention.

probability assessments, then—contrary to initial appearances—they are not necessarily going wrong in giving divergent responses in the help and harm conditions of our various experiments.<sup>11</sup> To test this alternative explanation of the epistemic side-effect effect, we presented modified versions of four of the vignette pairs described above to a new set of participants and asked them to rate the probabilities of the side-effects occurring.

## 5.1. Methods

### 5.1.1. Participants

845 undergraduate college students from a large, public university in the northeastern United States participated in this study on a voluntary basis. They were neither paid nor given course credit for their participation.

### 5.1.2. Materials and procedure

Participants were given truncated versions of the environment, movies, New Jersey and Nazi vignettes. The vignettes were shortened because each of the original vignettes includes information about how things turned out—e.g., “Sure enough, the environment was *helped/harmed*” or “Sure enough, the policy made the movies *better/worse* from an artistic standpoint.” To ensure that participant assessments of the likelihood of a side-effect occurring was not influenced by their knowledge that it actually did occur, this information was omitted. Except for ending with the chief protagonist saying “Let’s start the new program” or “Let’s implement the new policy,” the vignettes were otherwise exactly like the originals.

---

<sup>11</sup> The present suggestion may gain further support from the findings of Mele & Cushman (2007) that show that the objective probability of a result coming about has an important effect on attributions of intentionality—an effect that is greater than that of an agent’s subjective probability or credence.

Each participant was asked a question of the form “On a scale of 0 to 10, how likely do you think it is that [the protagonist’s primary action] will [lead to the good/bad side-effect]?” Participants were asked to circle a number on an 11-point scale, where ‘0’ was labeled ‘Highly Unlikely,’ ‘5’ as ‘Neither Likely nor Unlikely’ and ‘10’ as ‘Highly Likely.’

## 5.2. Results

Participant responses to the chairman and the environment case are consistent with the alternative hypothesis under consideration (cf. Figure 11).

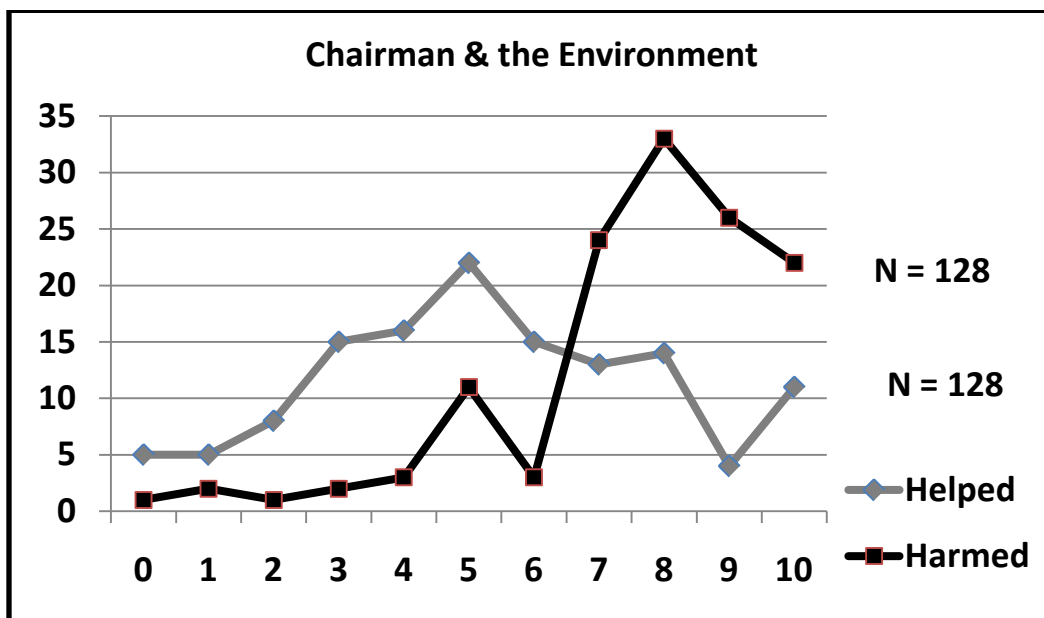


Figure 11

The mean response in the help condition is 5.30, and the median and mode responses are both 5. In other words, responses in the help condition tend to cluster around the neutral mid-point. In

the harm condition, however, the mean is 7.66, and the median and mode responses are 8. The difference between the two distributions is statically significant.<sup>12</sup>

Participant responses in the movie CEO case were also consistent with the target hypothesis (cf. Figure 12).<sup>13</sup>

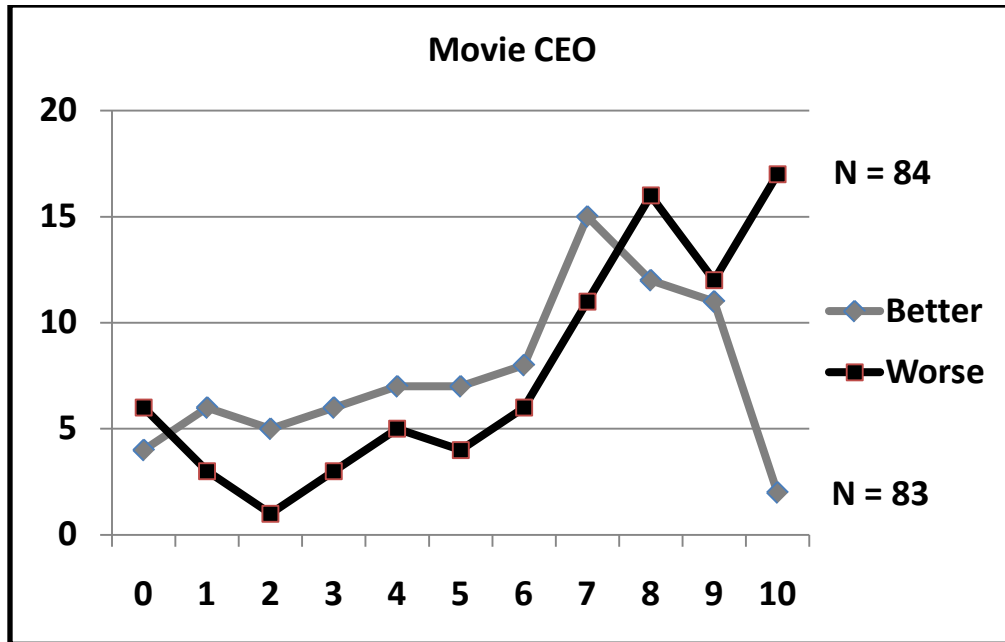


Figure 12

Means (better: 5.60, worse: 6.82), medians (better: 6, worse: 8), and modes (better: 7, worse: 10) all differ in the hypothesized direction.

In the increasing vs. decreasing sales in New Jersey case, however, there is no statistically significant difference between participant responses in the two conditions (cf. Figure 13).<sup>14</sup>

<sup>12</sup> T-test:  $t(240) = -8.029, p < 0.0001$ . Mann-Whitney U:  $U(256) = 12504.0, z = 7.28, p < 0.001$ .

<sup>13</sup> T-test:  $t(164) = -2.705, p < 0.01$ . Mann-Whitney U:  $U(167) = 4461.5, z = 3.12, p < 0.01$ .

<sup>14</sup> T-test:  $t(256) = -0.632, p > 0.05$ . Mann-Whitney U:  $U(258) = 8825.5, z = 0.86, p > 0.05$ .

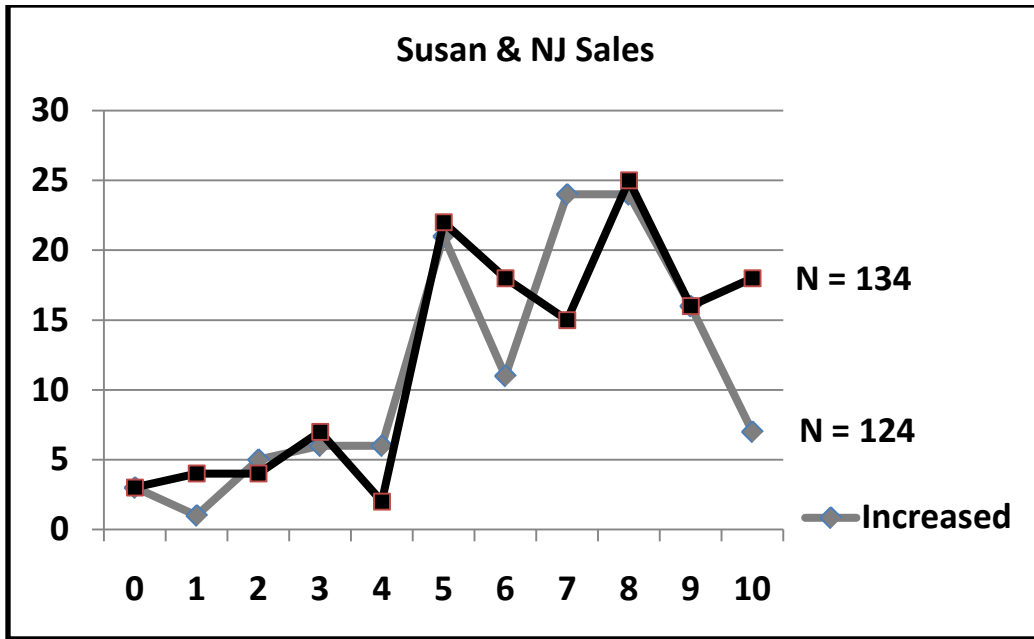


Figure 13

Neither the means (increase: 6.45, decrease: 6.63), medians (increase: 7, decrease: 7) nor modes (increase: 7 and 8, decrease: 8) indicated any appreciable but non-significant trend in the hypothesized direction.

In the case of the CEO living in Nazi Germany, the means (fulfill: 6.24, violate: 6.89) and medians (fulfill: 7, violate: 8) but not the modes (fulfill: 9, violate: 8 and 9) are all trending in the hypothesized direction, but the distributions of responses in the two conditions do not differ significantly (cf. Figure 14).<sup>15</sup>

<sup>15</sup> T-test:  $t(162) = -1.399, p > 0.05$ . Mann-Whitney U:  $U(164) = 3759.5, z = 1.32, p > 0.05$ .

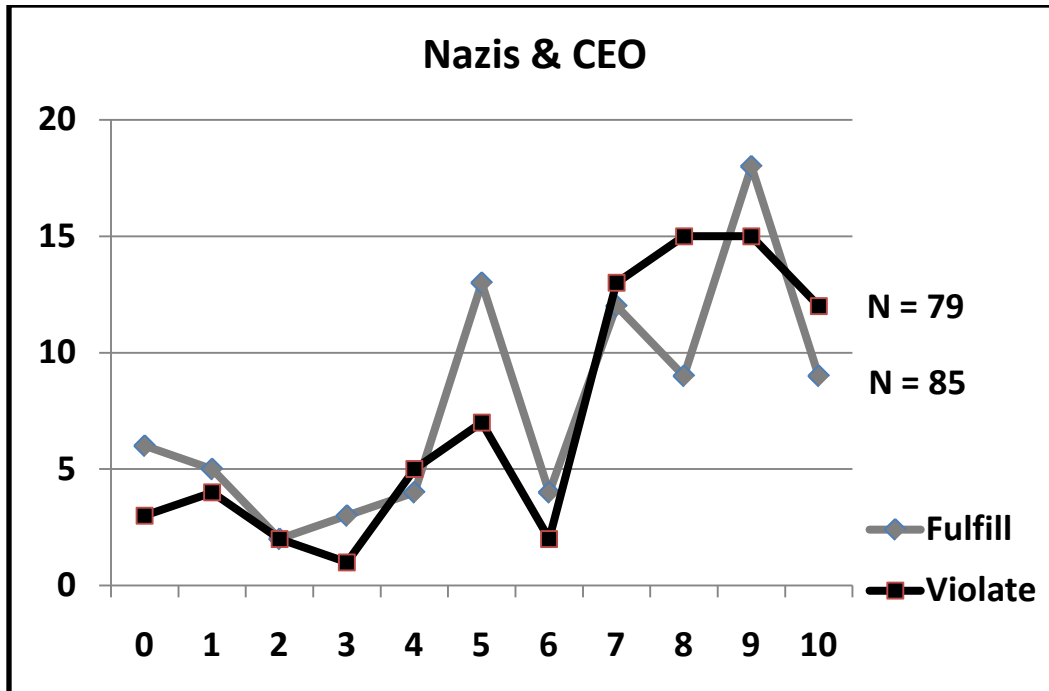


Figure 14

### 5.3. Discussion

According to the suggestion under consideration, the epistemic side-effect effect may be the result of participants recognizing (a) that the probability of succeeding in bringing about a harm is generally higher than the probability of succeeding in bringing about a help and (b) that the probability of *S* knowing that some future event will come about is generally higher than the probability of *S* knowing that some other future event will come about when, *ceteris paribus*, the probability of the first event's occurring is higher than the probability of the second. This hypothesis predicts that participants should give higher ratings when asked about the probability of an action leading to a harmful side-effect than when asked about an action leading to a helpful side-effect. In only one of out of the four cases investigated did we find a clear instance of the predicted pattern.

A more important point, however, is that fact that even if further investigation of participants' assessments of the probabilities of side-effects were to reveal the very pattern predicted by the hypothesis in question, this would not necessarily show that the hypothesis should be preferred to the hypothesis that the pattern of responses that constitute the epistemic side-effect effect is caused by participants responding to non-epistemic features of the research materials. The reason is that the participants' probability assessments may themselves be the result of them responding to non-epistemic features of the research materials. In other words, instead of participants first noting that harms are easier to bring about than helps and then making thoroughly rational judgments about protagonists' knowledge on this basis, it may be that the badness of the harms is leading participants not only to be more strongly inclined to attribute knowledge to protagonists but also to provide higher estimates of the probabilities of those harms. In order to test the foregoing suggestion, we conducted an experiment in which the two hypotheses under consideration would issue divergent predictions.

## **6. Experiment 5**

### 6.1. Methods

#### 6.1.1. Participants

376 undergraduate college students from a large, public university in the northeastern United States participated in this study on a voluntary basis. They were neither paid nor given course credit for their participation.

#### 6.1.2. Materials and procedure



Returning to the four cases involving the environment, the movie studio CEO, sales in New Jersey, and the CEO in Nazi Germany, we presented participants the same truncated versions of the help and harm vignettes used in Experiment 4. There was one added twist, however. Instead of having the testifying subordinate in the environmental harm condition say “We are sure that it will help us increase profits, but it will also harm the environment,” we had him say “We are sure that it will help us increase profits, but there is a *slight chance* that it will also harm the environment.” In the help condition the latter portion of this statement was changed to “and there is a *very strong chance* that it will also help the environment.” The subordinate in movie studio case reported, “There is also a *slight chance* that it will make our movies worse from an artistic standpoint” or “There is also a *very strong chance* that it will make our movies better from an artistic standpoint.” Susan’s subordinate reported, “It will simplify our corporate structure, but there is a *slight chance* that it will also decrease sales in New Jersey for the next quarter” or “It will simplify our corporate structure, and there is a *very strong chance* that it will also increase sales in New Jersey for the next quarter.” The subordinate of the CEO in Nazi Germany stated, “There is also a *slight chance* that you will be violating the requirements of the racial identification law” or “There is also a *very strong chance* that you will be fulfilling the requirements of the racial identification law.”

Participants were then asked to indicate the extent to which they agreed or disagreed with statements such as “The chairman knew that the new program would help the environment,” “The CEO knew that the new policy would make the movies worse from an artistic standpoint,” or “The CEO knew that the organizational changes would fulfill the requirements of the law.” As before, they marked their responses by circling a number on a seven-point scale, where ‘-3’ was labeled ‘Strongly Disagree,’ ‘0’ ‘Neutral,’ and ‘3’ ‘Strongly Agree.’ If participants are more

strongly influenced by explicit information about objective probabilities, we should find higher attributions of knowledge in each of the help conditions. However, if participants are more strongly influenced by the goodness or badness of the outcomes in question, we should find higher knowledge attributions in the harm conditions.

## 6.2. Results

In all four cases participants attributed knowledge to the chief protagonists in significantly greater numbers in the ‘slight chance harm’ conditions than in the ‘very strong chance help’ conditions. In the chairman and the environment case, the difference in distribution about the two means (help: 0.3, harm: 1.15) was statistically significant ( $t(93) = -2.064, p < 0.05$ ) (cf. Figure 15).<sup>16</sup> This difference was also reflected in the median (help: 0, harm: 2) and mode (help: -2, harm: 3) scores.

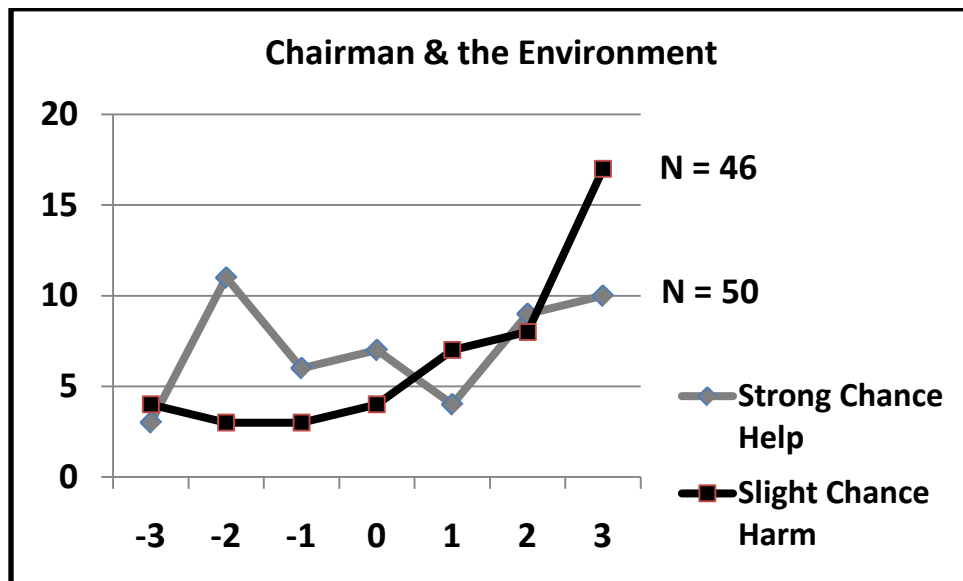


Figure 15

<sup>16</sup> A Mann-Whitney U test also indicates that the result was significant:  $U(96) = 1428.5, z = 2.04, p < 0.05$ .

Participants were also significantly more inclined to attribute knowledge to the movie studio CEO in the ‘slight chance worse’ condition than in the ‘very strong chance better’ condition (cf. Figure 16).<sup>17</sup> Other relevant statistics include the following: better mean = 0.31, worse mean = 1.27, better median = 0, worse median = 2, better mode = 2, worse mode = 3.

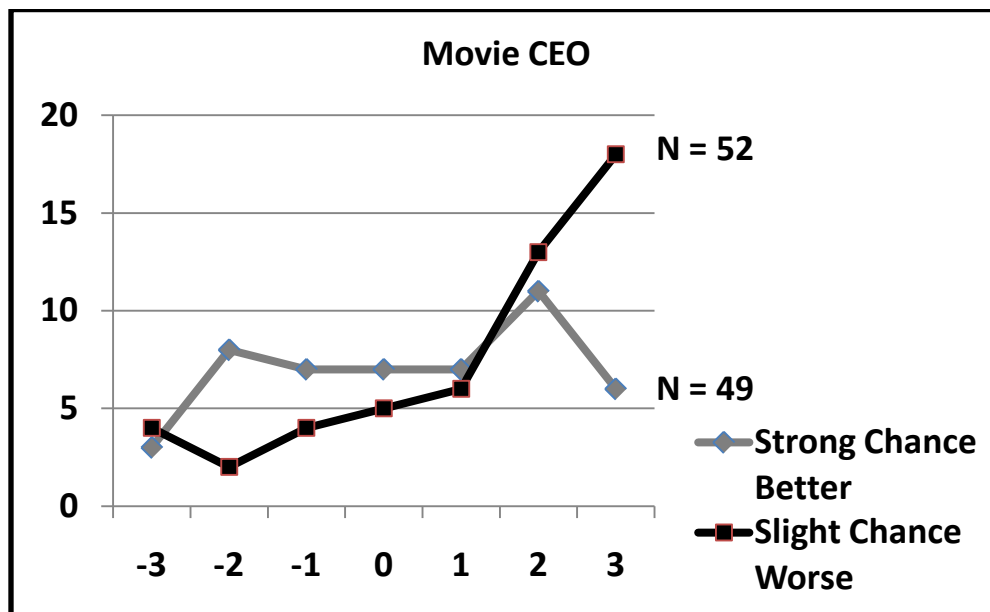


Figure 16

In the case of Susan increasing or decreasing sales in New Jersey, the means were 0.07 (increase) and 0.89 (decrease). The respective distributions were significantly different (cf. Figure 17).<sup>18</sup> Other relevant statistics include: increase median = 1, decrease median = 2, increase mode = 1, decrease mode = 2.

<sup>17</sup> T-test:  $t(99) = -2.572, p < 0.05$ . Mann-Whitney U:  $U(101) = 1672.0, z = 2.70, p < 0.01$ .

<sup>18</sup> T-test:  $t(88) = -2.213, p < 0.05$ . Mann-Whitney U:  $U(90) = 1293.5, z = 2.27, p < 0.05$ .

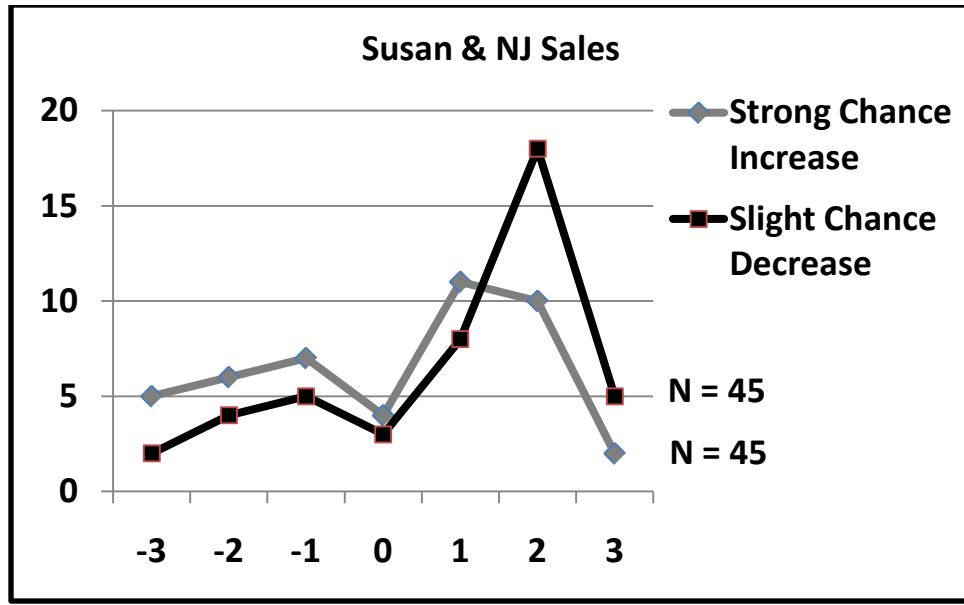


Figure 17

The distributions of participant responses in the fulfill and violate conditions of the Nazi and CEO case were significantly different (cf. Figure 18).<sup>19</sup> The means (fulfill: 0.4, violate: 1.61), medians (fulfill: 1, violate: 2) and modes (fulfill: 2, violate: 2 and 3) reflected this fact.

<sup>19</sup> T-test:  $t(75) = -3.371, p < 0.01$ . Mann-Whitney U:  $U(89) = 1322.5, z = 2.73, p < 0.01$ .

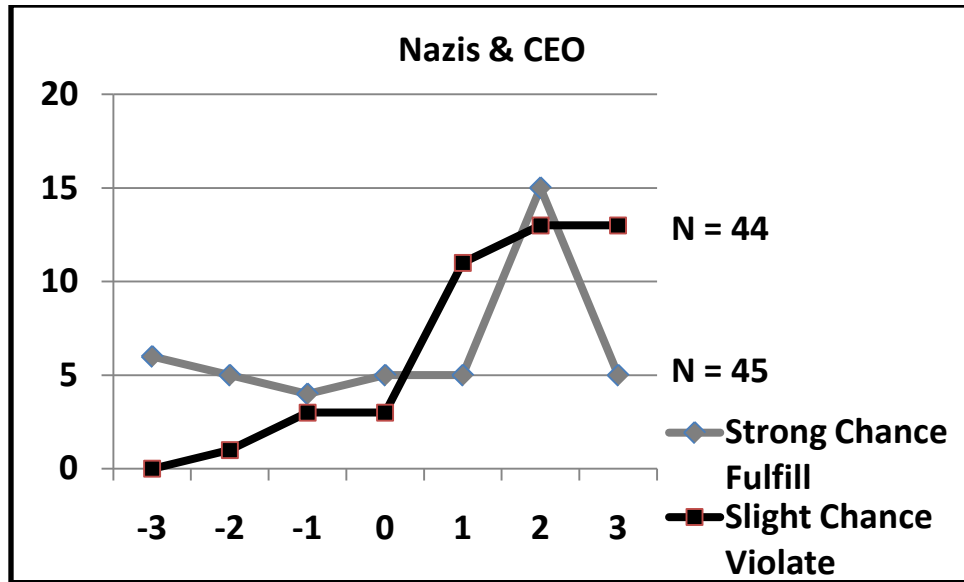


Figure 18

### 6.3. Discussion

In Experiment 2 we found that participants more readily attribute knowledge in harm cases than in help cases. In Experiment 4 we found that they ascribe greater probabilities to harmful side-effects than to helpful ones. One suggestion is that the latter phenomenon is responsible for explaining the former. Experiment 5, however, suggests that both phenomena have the same underlying cause. If participants in Beebe and Buckwalter’s study and in Experiment 2 were being guided primarily by their assessments of the probabilities of certain side-effects occurring, then in the present experiment they should be more strongly inclined to attribute knowledge when there is “a very strong chance” that a side-effect will occur than when there is only “a slight chance.” But this is not what we find. Rather, we find that participant attributions of knowledge are significantly more sensitive to the goodness or badness of the side-effects in question than to the objective information about probabilities they are explicitly given.

## 7. General discussion

Almost every scholar working on the Knobe effect has taken it to be glaringly obvious that the protagonists in both the help and the harm versions of the vignettes that appear in the literature genuinely know that the side-effects in question will occur. In Knobe's (2003a, p. 190) original experimental report, for example, he claims that the chairman "foresaw" the side-effect in both the help and the harm conditions. In later work Knobe (2006) claims that the chairman "decided to implement the program even though he specifically knew that he would thereby be harming the environment." (p. 206) Pettit and Knobe (2009) describe both help and harm versions as cases where "the agent knows that she will be bringing about a particular outcome through her behavior but that she does not care about this outcome in any way." (p. 587) Hugh McCann (2005) writes, "The situation of Knobe's chairman, however, is not one of negligence: in both vignettes, he knows perfectly well what he is doing." (p. 739) Frederick Adams (2006) writes, "the CEO knows he will help the environment (but does not care)... And, after all, [the chairman] was told that he would harm the environment. So he could not claim not to have known." (p. 260) Describing the sales in New Jersey case, Steven Sverdlik (2004) writes, "The example that Knobe and Mendlow use to exemplify this involves an executive who makes a decision knowing it will reduce sales in one area, but will also increase sales in another area by a greater amount." (p. 229) Nadelhoffer (2004b) writes, "Not surprisingly, to the extent that the CEO adopted the business plan knowing full well that it would decrease the sales in New Jersey, 75% of the subjects said that she intentionally brought this side effect about." (p. 264)

A great many scholars refer to the side-effects in the Knobe effect literature as "foreseen but undesired" (e.g., Knobe 2004a; Knobe 2006; Knobe & Burra 2006; Leslie, Knobe & Cohen

2006; Doris, Knobe & Woolfolk 2007; Knobe & Doris 2010; Nadelhoffer 2004a; Nadelhoffer 2006; Sverdlik 2004; McCann 2005; Mele 2006; Mele & Cushman 2007; Cushman & Mele 2007; Machery 2008; Hindriks 2008). Mele and Cushman (2007), for example, categorize each of the side-effect actions described in the cases above as ‘ASEAs’ or ‘anticipated side-effect actions.’ In calling the side-effect in these cases ‘anticipated,’ they are assuming that the actions are and will be taken by participants to be foreseen or foreknown by the agents in question. Our results show this assumption is false. As Beebe and Buckwalter (2010) note, some even take the chairman’s knowledge to be so obvious that they mistakenly think participants were explicitly told that that the chairman knew the side-effects would occur (e.g., Adams 2006, p. 259).

Several scholars go further and base their explanations of the Knobe effect in part upon the “fact” that protagonists in both help and harm conditions know that the side-effects will occur.<sup>20</sup> Nadelhoffer (2004b), for example, writes:

After all, in the case involving a CEO who did not care that the business plan he was adopting would help the environment, both of the conditions of the “desire plus knowledge” model were met—viz., the CEO wanted to adopted a business plan and he knew that adopting the business plan would help the environment—yet subjects gave the CEO a low intentionality rating (Knobe 2003a). This shows that desire and knowledge cannot be jointly sufficient for ascriptions of intentional action after all—at least as far as the majority of lay persons is concerned. (p. 267)

---

<sup>20</sup> Leslie, Knobe and Cohen (2006, 422) ask, “When [i.e., at what stage of human development] does the side-effect effect develop? One prerequisite is an appreciation that the actor knows the side effect will occur.” Being able to distinguish knowledge from non-knowledge may well be a prerequisite for displaying the side-effect effect, but it cannot be for the reasons these scholars suggest. It is not that children must recognize that in both the help and harm conditions the chairman knows the side-effect will occur. If a knowledge attribution capacity is a prerequisite, our results suggest that it will be because children will need to distinguish between the chairman knowing in the harm case but not knowing in the help case.

The results we report show that conclusions such as this are unwarranted because they are based upon mistaken and untested beliefs about how participants will view the protagonists' knowledge. Our results highlight the importance of experimentally testing all empirical claims scholars might wish to make about folk psychological thinking.

According to the received epistemological tradition, the following theses are both true:

( $\phi$ 1) Whether a true belief counts as knowledge depends only upon epistemic factors such as evidence or reliability.

( $\phi$ 2) Whether a true belief counts as knowledge is independent of whatever actions a believer may undertake in light of that belief.

Although taking a stand on the nature of the psychological processes involved in epistemic cognition is not essential to pursuing the central tasks of epistemology, it does seem that the following psychological theses have been widely (if only tacitly) accepted by epistemologists:

( $\psi$ 1) When deciding whether a true belief counts as knowledge, ordinary subjects will look only to epistemic factors such as evidence or reliability.

( $\psi$ 2) When deciding whether a true belief counts as knowledge, ordinary subjects will not take into account features of whatever actions a believer may undertake in light of that belief.

Our research on the epistemic side-effect effect clearly shows ( $\psi$ 2) to be false. We also think that our findings undermine ( $\psi$ 1), but the case for this claim is more tentative because it involves ruling out alternative explanations, as in Experiments 4 and 5. Whether our results call into question ( $\phi$ 1) or ( $\phi$ 2) is a more difficult question to answer because it depends upon a host of other issues concerning proper philosophical methodology. Experimental philosophers have



perhaps been too quick in trying to move straight from calling into question psychological theses such as ( $\psi$ 1) and ( $\psi$ 2) to calling into question central philosophical theses such as ( $\phi$ 1) and ( $\phi$ 2).

What is clear, however, is that a growing number of epistemologists have begun to reject ( $\phi$ 1), most notably those who defend some form of ‘pragmatic encroachment’ (e.g. Fantl & McGrath, 2002, 2007; Hawthorne, 2004; Hawthorne and Stanley, 2008; Stanley, 2005). These thinkers endorse the following, alternative thesis:

( $\phi$ 3) Whether a true belief counts as knowledge can depend in part upon non-epistemic facts about how much is at stake for a subject concerning the truth of the belief.

What is equally clear is that no epistemologist who rejects ( $\phi$ 1) would also want to endorse a thesis such as:

( $\phi$ 4) Whether a true belief counts as knowledge can depend in part upon non-epistemic facts about whether or not your belief leads you to make bad movies.

It has long seemed true that if two subjects have equally strong evidence for the same true belief,  $p$ , then either both subjects know that  $p$  or both fail to know that  $p$ . However, if more epistemologists begin to reject this idea and to allow room for non-epistemic factors to affect the truth of knowledge attributions, there will need to be principled ways to limit the range of permissible non-epistemic factors. Research on the epistemic side-effect effect presents a challenge to those proponents of pragmatic encroachment who wish to defend their views using what they claim are widely shared, intuitive responses to concrete cases (e.g., Hawthorne and Stanley). Our results show that the range of non-epistemic factors to which ordinary subjects are sensitive in making epistemic assessments is much larger and more varied than epistemologists have previously suspected. And most of them do not seem to be proper bases for distinguishing knowledge from non-knowledge.

Some issues that we will address in future research include the question of whether the degree of help or harm that is brought about by an agent can affect knowledge attributions. The experiments reported above only show a distinction between help and harm cases. But if the harm in one case is significantly greater than the harm in another, will this elicit even stronger, more confident knowledge attributions? Also, because the “harm” brought about by the CEO in Nazi Germany who violated an unjust law was not really harmful, further investigation of the relation between salient norm violations and knowledge attributions is warranted. Nadelhoffer (2004a) has argued that when protagonists are truly praiseworthy, intentional action attributions will match those in blameworthy cases. In order to see how closely the patterns of intentional action attributions in Knobe effect cases match the patterns of knowledge attributions in epistemic side-effect effect cases, it would be profitable to look into this matter. Finally, the difficult issue of the complex role that feelings of blame play in both intentional action and knowledge attributions is one that must be addressed without delay.<sup>21</sup>

Our results provide further evidence of how the practice of making epistemic evaluations is more intimately bound up with our more general practices of normative assessment than many philosophers have thought. They also provide additional confirmation for Knobe’s (2007, Pettit & Knobe, 2009) suggestion that the tight and complex set of relationships between folk psychology and normative assessment appears to leave no folk psychological concept untouched by our proclivity for evaluating the goodness or badness of all things.

---

<sup>21</sup> For detailed discussion of the issue of blame and intentional action attributions, cf. Nadelhoffer (2004b) and Phelan and Sarkissian (2008).

## References

- Adams, F. (2006). Intentions confer intentionality upon actions: A reply to Knobe and Burra. *Journal of Cognition and Culture*, 6, 132–46.
- Beebe, J. R., & Buckwalter, W. (2010). The epistemic side-effect effect. *Mind & Language*, 25, 474-498.
- Cullen, S. (2010). Survey-driven romanticism. *Review of Philosophy and Psychology*, 1, 275-96.
- Cushman, F., & Mele, A. (2007). “Intentional Action: Two-and-A-Half Folk Concepts?” In J. Knobe and S. Nichols (Eds.), *Experimental Philosophy* (pp. 171-188). New York: Oxford University Press.
- Doris, J., Knobe, J., & Woolfolk, R. L. (2007). Variantism about responsibility. *Philosophical Perspectives*, 21, 183-214.
- Fantl, J., & McGrath, M. (2002). Evidence, pragmatics, and justification. *Philosophical Review*, 111, 67-94.
- Fantl, J., & McGrath, M. (2007). On pragmatic encroachment in epistemology. *Philosophy and Phenomenological Research*, 75, 558-589.
- Guglielmo, S., & Malle, B. (forthcoming). Can unintended side effects be intentional? Resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin*.
- Hawthorne, J. (2004). *Knowledge and lotteries*. New York: Oxford.
- Hawthorne, J., & Stanley, J. (2008). Knowledge and action. *Journal of Philosophy*, 105, 571-590.

- Hindriks, F. (2008). Intentional action and the praise-blame asymmetry. *Philosophical Quarterly*, 58, 630-641.
- Knobe, J. (2003a). Intentional action and side effects in ordinary language. *Analysis*, 63, 190-3.
- Knobe, J. (2003b). Intentional action in folk psychology: an experimental investigation. *Philosophical Psychology*, 16, 309-24.
- Knobe, J. (2004a). Folk psychology and folk morality: Response to critics. *Journal of Theoretical and Philosophical Psychology*, 24, 270-279.
- Knobe, J. (2004b). Intention, intentional action and moral considerations. *Analysis*, 64, 181-187.
- Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies*, 130, 203-231.
- Knobe, J. (2007). Reason explanation in folk psychology. *Midwest Studies in Philosophy*, 31, 90-106.
- 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.
- Knobe, J., & Doris, J. (forthcoming). Strawsonian variations: Folk morality and the search for a unified theory. In J. Doris and the Moral Psychology Research Group (Eds.), *The Handbook of Moral Psychology*. Oxford: Oxford University Press.
- Knobe, J., & Mendlow G. (2004). The good, the bad, and the blameworthy: Understanding the role of evaluative considerations in folk psychology. *Journal of Theoretical and Philosophical Psychology*, 24, 252-258.
- 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.

- Machery, E. (2008). The folk concept of intentional action: Philosophical and experimental issues. *Mind & Language*, 23, 165-189.
- Malle, B. (2006). Intentionality, morality, and their relationship in human judgment. *Journal of Cognition and Culture*, 6, 87-112.
- Mallon, R. (forthcoming). Knobe vs. Machery: Testing the trade-off hypothesis. *Mind & Language*.
- McCann, H. (2005). Intentional action and intending: Recent empirical studies. *Philosophical Psychology*, 18, 737-48.
- Mele, A. (2006). The folk concept of intentional action: A commentary. *Journal of Cognition and Culture*, 6, 277-290.
- Mele, A., & Cushman, F. (2007). Intentional action, folk judgments, and stories: Sorting things out. *Midwest Studies in Philosophy*, 31: 184-201.
- Nadelhoffer, T. (2004a). On praise, side effects, and folk ascriptions of intentionality. *Journal of Theoretical and Philosophical Psychology*, 24, 196-213.
- Nadelhoffer, T. (2004b). Blame, badness, and intentional action: A reply to Knobe and Mendlow. *Journal of Theoretical and Philosophical Psychology*, 24, 259-269.
- Nadelhoffer, T. (2006). Bad acts, blameworthy agents, and intentional action: Some problems for juror impartiality. *Philosophical Explorations*, 9, 203-219.
- Pettit, D., & Knobe, J. (2009). The pervasive impact of moral judgment. *Mind & Language*, 24, 586-604.
- Phelan, M., & Sarkissian, H. (2008). The folk strike back: Or, why you didn't do it intentionally, though it was bad and you knew it. *Philosophical Studies*, 138, 291-298.

- Schwarz, N. (1995). What respondents learn from questionnaires: The survey interview and the logic of conversation. *International Statistical Review*, 63, 153-168.
- Stanley, J. (2005). *Knowledge and practical interests*. New York: Oxford University Press.
- Sverdlik, S. (2004). Intentionality and moral judgments in commonsense thought about action. *Journal of Theoretical and Philosophical Psychology*, 24, 224-236.
- Tannenbaum, D., Ditto, P.H. & Pizarro, D.A. (2007). *Different moral values produce different judgments of intentional action*. Unpublished manuscript, University of California-Irvine.