Rev.Phil.Psych. (2013) 4:639–657
DOI 10.1007/s13164-013-0156-1

Is What is Worse More Likely?—The Probabilistic Explanation of the Epistemic Side-Effect Effect

Nikolaus Dalbauer · Andreas Hergovich

Published online: 19 July 2013

Abstract One aim of this article is to explore the connection between the Knobe effect and the epistemic side-effect effect (ESEE). Additionally, we report evidence about a further generalization regarding probability judgments. We demonstrate that all effects can be found within German material, using 'absichtlich' [intentionally], 'wissen' [know] and 'wahrscheinlich' [likely]. As the explanations discussed with regard to the Knobe effect do not suffice to explicate the ESEE, we survey whether the characteristic asymmetry in knowledge judgments is caused by a differing perception of probabilities concerning the occurrence of the side-effects. Our findings show that a negative side-effect is judged more probable, even if the objective probabilities would suggest otherwise. We argue that the best explanation for these results is that the Knobe effect applies to the perception of probabilities as well: a probabilistic side-effect effect.

1 Introduction

In recent years, experimental philosophers have demonstrated that many concepts important to philosophical theorizing are prone to be influenced by factors that have never been considered. One of the most significant findings in this regard is called the Knobe effect. Knobe (2003) and others have demonstrated that we judge an action as intentional if it brings about a negative side-effect, but not if the outcome is positive. This asymmetry could not only be found in cases of intentionality ratings. Beebe and Buckwalter (2010) have demonstrated that people are also more willing to ascribe knowledge about the outcome of an action to a person if it includes a negative side-effect. This effect is thought to be another instance of the Knobe effect, although recently discussed explanations of that effect do not apply to the new case. Therefore,

N. Dalbauer (✉)
Department of Applied Psychology: Applied Social Psychology and Consumer Research, University of Vienna, Universitätsstraße 7/C0717, 1010 Vienna, Austria
e-mail: nikolaus.dalbauer@gmx.at

A. Hergovich
Department of Applied Psychology: Applied Social Psychology and Consumer Research, University of Vienna, Universitätsstraße 7/B0613, 1010 Vienna, Austria

Beebe and Jensen (2012) propose that the epistemic side-effect effect might be caused by probability ratings.

In Section 2 we introduce the Knobe effect and summarize some of the most important findings in this field of research. Section 3 deals with the connection between the Knobe effect and the ESEE. For instance, previous experiments suggest that gender effects exist for the ESEE but not for the Knobe effect. Furthermore, we discuss the way in which Beebe and Jensen (2012) have tried to explore the influence of probability ratings on knowledge judgments. In Experiment 1 (Section 4) we try to remediate the shortcomings of their work. The results suggest that the probability judgments themselves seem to be influenced by the valence of a side-effect. Therefore, we ran an additional survey to gain more evidence to support these findings (Section 5). Finally (Section 6), we discuss the results and finish by concluding that the Knobe effect applies to the perception of probabilities as well.

2 Researching the Knobe Effect

As we go through moral development, it is an important step to learn that there might be a disparity between the intention of an acting person and the result of that action (cf. Lane et al. 2010). At some point, children realize that an action may be performed intentionally or unintentionally. Beginning with this moment of understanding, this distinction will affect the way they treat other people and, consequently, they themselves will be judged along those lines. Yet, such evaluations may lead to severe consequences. Just think about a court trial and the great difference it could make whether the offence was performed intentionally or not. Due to the significant role that ascriptions of intentional action play in everyday life, it is not astounding that the concept of intentional action is of great importance for action theory literature (cf. Mele 1992, 1997).

To explore the folk concept of intentionality, Joshua Knobe developed the following hypothetical case with two different conditions:

> 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 [and] it will also harm [help] the environment.' The chairman of the board answered, 'I don't care at all about harming [helping] 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 [helped]. (Knobe 2003, p.191)

If the outcome of an action is foreseen but presents neither the purpose of that action nor a means to that goal, then that consequence is a side-effect. By this definition, harming or helping the environment is a side-effect and, as may be concluded from the definition, a side-effect action should be performed unintentionally. But would we say that the chairman harmed (helped) the environment unintentionally?

Knobe (2003) carried out a survey and simply asked everyday speakers this question. The result was that 82 % of the participants declared that the chairman intentionally harmed the environment, but only 23 % thought that the chairman intentionally helped

the environment. Knobe concluded "that people's intuitions as to whether or not a behavior was performed intentionally can be influenced by their beliefs about the moral status of the behavior itself" (2004a, p.270).

As these findings have been able to be replicated (Adams and Steadman 2004a, b; Knobe 2004b; Knobe and Mendlow 2004; Nadelhoffer 2004), they initiated a vivid discussion about the source of this asymmetry, which was called the Knobe effect, the side-effect effect, or just SEE (cf. Feltz 2007). Some authors argued that this effect is due to a tendency of people to blame someone who foresees that her action will be harmful but executes it nevertheless (Adams and Steadman 2004a, b). Others think that there are different folk concepts of intentionality (Cushman and Mele 2008; Lanteri 2012). Yet others explain the asymmetry by a deep self-concordance, which states that people judge an action as intentional if it goes along with this person's basic beliefs and attitudes, that being the best way of predicting her actions in the future (Sripada 2010; Sripada and Konrath 2011). Even though many approaches seem very reasonable, they have not proven resistant to falsification through counterexamples, which has caused severe doubts about their validity up to now. Therefore, one may agree with the analysis of Nichols and Ulatowski (2007, p.346): "What has not yet emerged, though, is a satisfying explanation of the phenomenon. In our view, the literature makes apparent that all of the prominent explanations have serious shortcomings. The puzzle persists" (cf. Knobe et al. 2012).

Simultaneously, researchers began to explore the pervasiveness of the SEE. They found that this effect can be demonstrated with children as young as 4 to 5 years (Leslie et al. 2006; Michelin et al. 2010; Pellizzoni et al. 2009). It proves to be independent of age, gender, religion, educational level, a potential major in moral philosophy, and English as the first language of the participants (Young et al. 2006). The side-effect of the program does not have to be morally bad to cause a response asymmetry. Through the use of various vignettes, it could be demonstrated that the SEE also appears when aesthetic or economic norms are violated (Knobe 2004a; Knobe and Mendlow 2004). Uttich and Lombrozo (2010) have even shown that in cases where more than one norm is important, the most salient norm triggers the direction of the SEE. If a story contains conflicting norms, moral considerations are not necessarily the most important ones. Accordingly, in such scenarios the SEE is not dependent on the moral status of the action.

A hotly debated field in experimental philosophy concerns intercultural differences in philosophical intuitions. In this respect, it is of general interest to examine whether the Knobe effect varies between different countries' respective languages. The claim of cross-cultural disparity within intuition is grounded in two famous studies (Machery et al. 2004; Weinberg et al. 2001) and much additional theoretical work (e.g., Nichols et al. 2003; Mallon et al. 2009; Machery et al. 2012). Considering the severe critique that challenges the validity of these studies both theoretically (Cullen 2010; Deutsch 2009; Marti 2009) and empirically (Machery et al. 2009; Nagel 2012; Sytsma and Livengood 2011), it is unclear to what extent intercultural differences really do exist. Yet, there is evidence that at least the Knobe effect is culturally universal. Knobe and Burra (2006) replicated the characteristic asymmetry using a Hindi translation of the original chairman vignette. Michelin et al. (2010) and Pellizzoni et al. (2009) also found the Knobe effect using Italian material. However, until now, no one has studied this effect in German.

Another area of research substitutes the word "intentionally" in order to see how robust the SEE is. In this respect, a significant albeit smaller asymmetry could be found when the participants were asked: "Did the chairman intend to harm [help] the environment?" and "Was it the chairman's intention to harm [help] the environment?" (Knobe 2004b; McCann 2005). Further studies have shown the same kind of effect with "decide," "want," "desire," "advocate," "in favor of," and, by negating the question, also with "oppose" (Pettit and Knobe 2009; Guglielmo and Malle 2010).

Although the pervasiveness of the Knobe effect is an interesting fact, the meanings of all the different questions somehow add up to a similar concept, sharing the core of an attitude in favour of harming (or helping) the environment regardless of the exact word choice. Therefore, it can be assumed that all of these effects were caused by the same mechanisms. These generalizations show that an explanation of the Knobe effect should not concentrate on the meaning of a single word such as "intentional." Nevertheless, a unified account may still be given, if it manages to explain why the attitude of the chairman is judged differently (cf. Pettit and Knobe 2009).

3 Surprising Generalization – The Epistemic Side-Effect Effect

Beebe and Buckwalter (2010) explored something different, though closely related. They were not interested in the ascription of intentionality, but in the attribution of knowledge. Therefore, they asked: "Did the chairman know that the new program would help/harm the environment?" (p.476). Note that the epistemic conditions for the chairman were exactly the same in both vignettes. A scale ranging from −3 ("the chairman didn't know") to 3 ("the chairman knew") was used to help participants express their opinion. Again, a difference was observed. The mean in the harm condition was 2.25, the average in the help condition 0.91. Beebe and Buckwalter have named this asymmetry within knowledge attribution the *epistemic side-effect effect* (ESEE). Despite the pervasiveness reported above, this result is quite surprising. On the one hand, to date no explanation would have permitted this effect to be foreseen. On the other hand, as Beebe and Buckwalter (p.447) have already stated: "Many scholars assume that the chairman knows full well that his actions will have the side-effects described." They were the first to consider that people could think differently. Furthermore, this finding cannot be understood as an attitude in favor of harming the environment.

Therefore, one might wonder whether this finding really is a generalization of the Knobe effect that shares a common cause, or if it actually has nothing to do with it. Yet, there is a reason for thinking that the ESEE does not have the same source as the SEE: Buckwalter and Stich (2013) report a significant gender effect when reanalyzing the data from Beebe and Buckwalter (2010). Seemingly, women show a bigger asymmetry in their knowledge ratings than men. In this context, it is important to remember that Young et al. (2006) did not find any gender differences concerning the SEE. These results are only contradictory if the ESEE is perceived as a generalization of the SEE. But if the ESEE has an independent source, then there is nothing puzzling about finding a gender effect for the ESEE but none for the SEE.

In any case, it seems important to test whether there are gender effects concerning the SEE and the ESEE using data from the same subjects, in order to determine whether there really is such a difference between these two effects.

The possibility that the ESEE and the Knobe effect do not share a common cause permits consideration of an explanation that focuses only on the asymmetry concerning knowledge ascription. In this regard, Beebe and Jensen (2012, p.690) asked: "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?" From the definition of knowledge, it may be inferred that one cannot know that something will happen if it does not actually happen in the end. So, if something is unlikely to happen, then it is also unlikely that someone can know that it might happen, unless she has strong evidence in favor of it. Therefore, to gain knowledge of an unlikely event, a deep understanding of the mechanisms in question is needed. It is quite possible that the statement of the vice-president does not suffice to convince participants that the environment will indeed be helped. In this regard, it is not surprising that the subjects of Beebe and Buckwalter's study ascribe a lesser degree of knowledge in the help condition, if they think that this condition is less likely, than in the harm condition.

Beebe and Jensen (2012) took up this line of thought and tried to put it to the test. In a series of experiments, they not only managed to replicate the ESEE and demonstrate that an asymmetry concerning knowledge attribution can be found in a great variety of those cases that had already been studied to explore the Knobe effect, but they also measured the conceivable probability of the side-effect. This confirmed that people think it is much more likely to harm the environment than to help it. However, that does not prove that the differences in probability ratings really are the cause of the ESEE. To test this explanation, Beebe and Jensen modified the vignettes in order to change the likelihood of the side-effects. In their new harm condition, the vice-president said about the new program: "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 he reported: "We are sure that it will help us increase profits and there is a *very strong chance* that it will also help the environment" (Beebe and Jensen 2012, p.707). By means of this modification, the help condition was intended to become more likely than the harm condition. Therefore, Beebe and Jensen expected that in this case people would be more strongly inclined to ascribe knowledge to the outcome of the positive side-effect. Nevertheless, the opposite turned out to be true. As before, the mean rating relating to knowledge in the help condition was 0.3 and in the harm version 1.15, on a scale ranging from −3 to 3. This asymmetry was statistically significant, which was construed in such a way that the probability of the outcome of an action was not the source of the ESEE.

One possible point of criticism concerns Beebe and Jensen's way of falsifying this explanation. They did not demonstrate that the subjects now actually perceived the side-effect in the help condition as more likely than in the harm condition. Although the respondents read that there is only "a slight chance" of harming the environment, we do not know how this statement was actually interpreted. We should not forget that the subjects in Beebe and Jensen's studies read this kind of vignette for the first time and did not compare it to the original story. So the simple fact of mentioning the possibility of harming the environment may have been sufficient to convince participants that the environment will be harmed eventually; especially if they believe that what increases a big company's profits is bad for the environment by definition.

Likewise, it is hard to see why people should perceive the positive side-effect as more likely in the modified version than in the original one. Quite the contrary: to read that there is a "very strong chance" that something will happen is less convincing than to read only that it will happen. Furthermore, in both conditions the specifications of the probabilities might have been interpreted as an attempt by the vice-president to persuade the chairman to permit the project. So it appears quite possible that the side-effect is still perceived as more probable in the harm condition than in the help condition.

In that regard, it seems to be necessary to examine the influence of the probability as perceived by the respondents. Therefore, we tested the probabilistic explanation in a slightly different way. Instead of manipulating the likelihood of the side-effect by using a direct description, we chose a scenario in which a certain positive side-effect would be more likely than a negative one. Only if the perceived probabilities really vary in a controlled way can it be possible to figure out whether they are a cause of the ESEE. If so, that would actually lead to the conclusion that this effect is not a generalization of the Knobe effect at all.

Experiment 1 aims to find out whether the Knobe effect and the ESEE can be observed within a German-speaking population, using "absichtlich" and "wissen" as the German translation of "intentionally" and "know." Furthermore, we want to replicate the different results concerning gender effects, which would be hinting at the nonexistence of a common cause. However, the main target of Experiment 1 is to analyze the connection between the perceived likelihood of a side-effect and the readiness to say that someone knew that something would happen.

4 Experiment 1

4.1 Participants

269 participants (72.5 % women, 27.5 % men between the ages of 15–70 years[1]) were tested online from 05-15-2011 to 06-14-2011 via www.soscisurvey.com. They were solicited by mail and bulletin boards to participate voluntarily in the study.

4.2 Materials and Procedure

In a between-subjects design, each participant was randomly assigned to one of four different versions of a questionnaire, each containing two vignettes. (The order of these vignettes was counterbalanced, as well as the valence of the side-effect; i.e., one half of the participants read about a negative side-effect and the other half about a positive one.) Furthermore, the online questionnaire was programmed to balance the proportion of men and women in each condition.

One of the vignettes was simply a German version of Knobe's original chairman story. We named this vignette the "profit story," as the purpose of the new program is to increase the company's profit. Secondly, we used a new variation of the chairman story, the "guideline story." In this story, the purpose of the new program is not to

[1] Mean=28.66 SD=9.92.

increase profits, but to conform to new and more restrictive environmental protection guidelines. Translated into English it reads 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 conform to the new and more restrictive environmental protection guidelines, but [and] it will also harm [help] the environment.' The chairman of the board answered, 'I don't care at all about harming [helping] the environment. I just want to abide by these guidelines. Let's start the new program.' Thus, they started the new program. Sure enough, the environment was harmed [helped].

The participants were instructed to read each story and specify their level of agreement with statements regarding the chairman's knowledge ("The chairman knew that the new program was going to harm [help] the environment.") and the chairman's intentionality ("The chairman harmed [helped] the environment intentionally.") using a visual analog scale ranging from "complete disagreement" to "complete agreement." The position opted for on the scale was automatically transferred into a value between 1 and 100. As it is usually the purpose of a new and more restrictive environmental guideline to help the environment and in the guideline story it is the very purpose of the new program to fulfill such a guideline, we expected that in this scenario a positive side-effect would be more likely than a negative one. To verify this assumption we used two questions about the probability of the side-effects: a more general and a specific one. The general probability question was: "How likely is it, that a new program which is used to increase profits [to conform to new environmental guidelines] will harm the environment?" With this question we focused merely on the harm a new program could cause, in order to be able to compare the answers of all respondents independently of the version of the given questionnaire. Therefore, this question always asked about the likelihood of harming the environment. In contrast, the specific probability question was a statement about the likelihood of the scenario described in the particular vignette that the participants had just read: "It was likely, that this program would harm [help] the environment." Participants were asked to rate both questions using a visual analog scale ranging from "unlikely" to "likely."

4.3 Results

As a first step, we looked for order effects concerning the ratings of intentionality, knowledge, and probability. There were no significant results; therefore we combined the two versions of the questionnaire in varying orders.

Subsequently, we analyzed the answers to Knobe's original chairman story to see whether the SEE and the ESEE could be replicated. In the help condition the mean for the intentionality rating was 11.64 (SD=20.0). In the harm condition it was 69.46 (SD=31.94). This difference was highly significant. There was also a significant difference concerning knowledge ascription. The average rating in the help condition was 87.05 (SD=26.65) but 96.07 (SD=15.71) in the harm condition (see Fig. 1).[2]

[2] Welch's (1947) *t*-test: Intentionality: $t(236.17)=-17.94$, $p<0.001$, d=2.33. Knowledge: $t(203.88)=-3.35$, $p<0.001$, d=0.47.

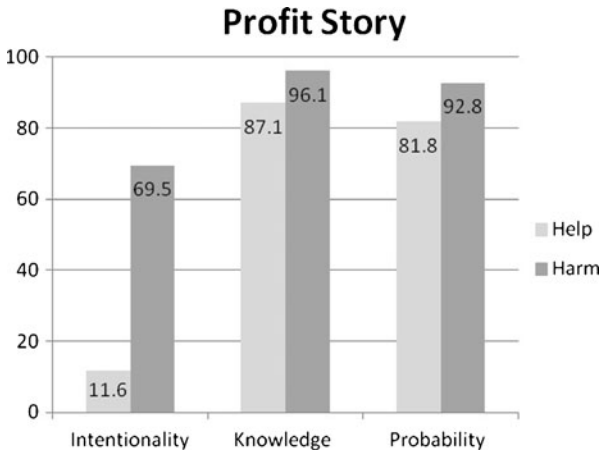

Fig. 1 Responses in all conditions of the profit story

Next, we examined our assumption that a program developed for the increase of profits was more likely to harm the environment than one meant to implement a new and stricter environmental guideline. The corresponding question was: "How likely is it that a new program used to increase profit [to conform to new preservation guidelines] will harm the environment?" The mean rating for profit was 60.3 (SD=27.83), but in the case that a new environmental guideline was the purpose it was 28.0 (SD=25.25). This difference was significant again and confirmed that our story about the different purpose of the new program could indeed change the perceived probability ratings (see Fig. 2).[3]

In a further step, we checked whether we could find the same ESEE for our guideline story. The average knowledge rating in the harm condition was 95.39 (SD=17.86) and the average for help was 87.24 (SD=24.07). Again, this difference was significant and resembled the results of the original chairman story. This was also true for the intentionality ratings, with 15.61 (SD=24.37) in the help condition and 58.56 (SD=32.38) in the harm condition (see Fig. 3).[4]

In addition to the general question about the probability of the side-effect, we also analyzed the "specific probability question," which asked how likely the outcome of the new program was as described in the story. For Knobe's original chairman story the mean in the help condition was 81.77 (SD=21.78) and 92.8 (SD=15.56) in the harm condition. The difference was expected and statistically significant. Surprisingly, and contrary to the results retrieved from the general probability question that we asked in order to check our assumptions about the new story, the results concerning the specific probability question in the guideline story did not vary from the results of the profit story. The average probability rating in the help condition was 85.02 (SD=21.4) and 91.74 (SD=14.75) in the harm condition. Again, this difference was significant.[5] (The results of the specific probability question are also illustrated in Figs. 1 and 3.)

[3] $t(268)=-15.19$, $p<0.001$, d=1.86.

[4] Welch's t-test : Intentionality: $t(256.96)=-12.95$, $p<0.001$, d=1.62. Knowledge: $t(235.11)=-3.13$, $p<0.005$, d=0.41.

[5] Welch's t-test: Profit story: $t(229,83)=-4,75$, $p<0.001$, d=0.63. Guideline story: $t(224.97)=-2.974$, $p<0.005$, d=0.40.

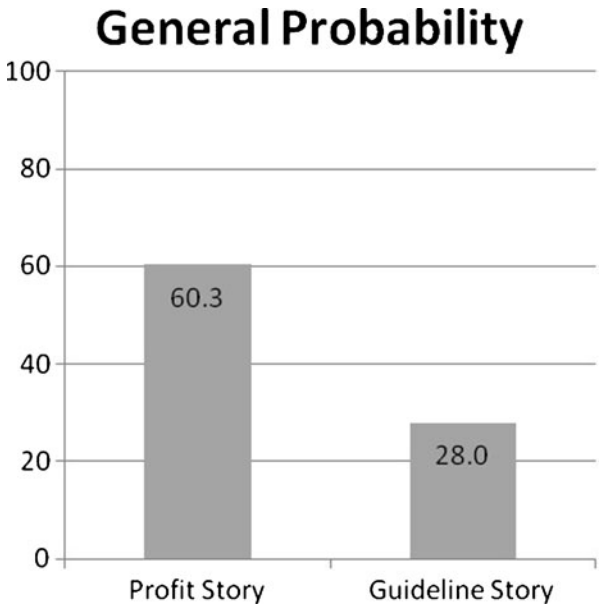

Fig. 2 Responses to the general probability question

There were also significant correlations between knowledge attribution and specific probability ratings, with r=0.33 (p<0.001) in the profit story and r=0.32 (p<0.001) in the guideline story. In order to test whether these correlations with medium effect sizes could be responsible for the ESEE, we conducted a mediation analysis. We tested whether the effect of valence on knowledge ratings was significantly influenced by probability perception, and whether the direct effect of valence on knowledge attribution would still be significant after controlling for this potential mediational effect. We used the Preacher and Hayes (2008) SPSS macro to calculate the total, direct, and indirect effect in both stories and found all effects to be significant. As suggested by Hayes (2009), we calculated bias-corrected bootstrap

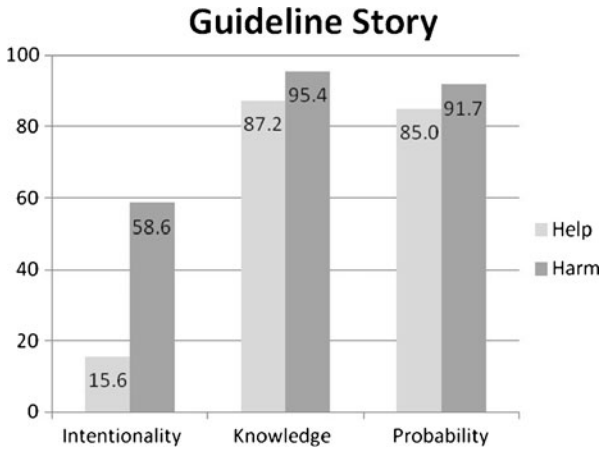

Fig. 3 Responses in all conditions of the guideline story

confidence intervals based on 5,000 bootstrap samples to estimate the statistical significance of the indirect effect. This effect was significant too.[6]

Finally, we examined whether the results differed between women and men. A 2×2 (gender x condition) between-subjects ANOVA was performed for each of the three dependent variables in both stories. There were no significant main or interaction effects of gender.[7]

4.4 Discussion

The results of Experiment 1 show that both the SEE and the ESEE can be found using German material. This means that the findings from Knobe (2003) with respect to Beebe and Buckwalter (2010) hold true for an Austrian sample as well. Together with the work from Knobe and Burra (2006) and Pellizzoni et al. (2009), there is strong evidence that the Knobe effect and the ESEE are culturally universal. Furthermore, these effects do not only appear with Knobe's original chairman story, as illustrated by the responses to our guideline story. Considering the large amount of existing literature, our experiment can be seen as another confirmation of the pervasiveness of the Knobe effect and the ESEE. Nevertheless, it must be noted that the ESEE is considerably smaller than the original Knobe effect. There were no gender-specific differences for these effects in either story.

We created the guideline story in order to present a story in which the positive side-effect appears to be more likely than the negative one. We hypothesized that if the ESEE were caused by probability judgments, the results regarding this story should show an inverted pattern, due to the assumption that people would more willingly attribute knowledge to the protagonist in the case of the positive side-effect (help condition). However, this did not turn out the way we had expected. The characteristic response pattern of the ESEE was the same for both stories in our survey.

There are two possible ways to explain these findings: either the ascription of knowledge does not depend on the perceived probabilities, or such an effect really exists but the side-effect in the guideline story was not judged as more likely in the help condition. It is hard to tell which of these two possibilities actually holds true, because the results do not fully support either one of them. This is due to the two different questions we used to measure the perception of probability. One question

[6] Profit story: total effect: $t(268)=3.41$, $p<0.001$, direct effect: $t(267)=2$, $p<0.05$ indirect effect: 99 % C.I.: 0.87 – 8.82.
Guideline story: total effect: $t(268)=3.17$ $p<0.01$ direct effect: $t(267)=2.49$, $p<0.05$ indirect effect: 99 % C.I.=0.09 – 5.77.

[7] Profit story:
Main effect: intentionality: $F(1,269)=0.14$, $p>0.05$; knowledge: $F=0.002$, $p>0.05$; specific probability: $F=0.87$, $p>0.05$.
Interaction: intentionality: $F(1,269)=1.5$, $p>0.05$; knowledge: $F=1.01$, $p>0.05$; specific probability: $F=0.05$, $p>0.05$.
Guideline story:
Main effect: intentionality: $F(1,269)=0.4$, $p>0.05$; knowledge: $F=0.52$, $p>0.05$; specific probability: $F=0.35$, $p>0.05$.
Interaction: intentionality: $F(1,269)=0.09$, $p>0.05$; knowledge: $F=0.72$, $p>0.05$; specific probability: $F=4$, $p>0.05$.

merely asked in general how likely it is that a program will harm the environment, depending on the purpose of the program. The answers to this question show that participants rated the probability that a program developed to meet the new guidelines for environmental preservation would harm the environment as far lower than if this program served the increase of profits. With a mean above the midpoint (60.3), people think that it is in fact likely that making profit harms the environment; with a mean below the midpoint (28.0), people think that this consequence is unlikely if the purpose of the program is meeting the preservation guidelines.

The second question was posed directly after the subjects had read the story and referred to it. Here, the participants rated the likelihood of the particular ending of the story. While the response to the general probability question shows that the story type makes a difference to the probability rating, this could not be confirmed by the specific probability question. The answers to this question only varied as a function of the outcome factor (positive vs. negative side-effect). People estimate the harm outcome as more likely than the help outcome for both stories. Additionally, there is a relationship between the specific probability and the knowledge ratings, as the correlations show.

Furthermore, the mediation analyses show a significant indirect effect, which supports the thesis that the probability of an event could have an impact on the ascription of knowledge about this event. Nevertheless, the mediation analyses also show that there is still a significant asymmetry in knowledge ratings. Therefore, we can conclude that the ESEE is not caused by a differing likelihood of the side-effects described.

So it seems that although the guideline story describes a program that is generally not associated with harming the environment, the way in which this program is presented within the story leads participants to judge the harm condition as more likely than the help condition. An explanation of these results could be that probability ratings are influenced by the valence of the side-effect in the same way that knowledge attributions are. This could also explain why the perception of probability changes, depending on whether the question is asked generally or specifically. This finding may represent a further generalization of the Knobe effect: a probabilistic side-effect effect (PSEE).

A similar thought has already been suggested by Egré (2010, in press), who argues that probability judgements are susceptible to normative considerations in the same way judgements about intentionality are.[8] In this context, Egré cites a study from Pighin et al. (2011), which shows that a specific numeric probability for a particular disease is perceived differently depending on the severity of that disease. This result can easily be seen as another instance of the PSEE.

Nevertheless, one might object that perhaps the differences in probability ratings found in Experiment 1 occur only because of certain attitudes that people have towards big companies and their influence on the environment. Perhaps it is always more likely for a big company to harm the environment and that is all there is to the PSEE.

Although the answers to the general probability question suggest that people do differentiate between different methods and their influence on the environment, we

[8] Unfortunately Egré (2013) broad account cannot be elaborated here.

think that Experiment 1 gives only preliminary evidence in favor of a PSEE. To explore this effect further, we designed Experiment 2, which provides another two stories.

5 Experiment 2

5.1 Participants

340 people (56.8 % women, 43.2 % men between the ages of 11–77 years[9]) voluntarily participated in Experiment 2. Again, they were surveyed by mail and bulletin boards.

5.2 Materials and Procedure

In a between-subjects design, each participant was randomly assigned to either the positive or the negative condition of the two stories. Again, the online questionnaire was programmed to balance the proportion of men and women in each condition. In counterbalanced order we used German translations of a vignette first suggested by Turner (2004) and applied by Knobe (2004a) – the "movie story" – and one from Knobe and Mendlow (2004) – the "reorganization story." In Experiment 1, the main purpose of the action under consideration was alternated but the side-effect stayed the same. Therefore, in Experiment 2 we decided to alternate the side-effect to examine whether the PSEE might simply be caused by attitudes towards the environment (and its protection).

Beebe and Jensen (2012) have already demonstrated that an ESEE can be found using both stories. We translated their version of the movie story, as the original one contains a statement about the knowledge of the chairman.

> 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. (Beebe and Jensen 2012, p.695)

In this story, the CEO of a big movie company wants to increase profits and harms [or helps] the quality of the movies produced as a side-effect. It can be argued that better movies will lead to a higher number of people paying to see them and subsequently to more profits. Therefore, it seems highly plausible that a program intended to maximize profits will enhance the quality of the movies. Compared to the profit story, where caring about the environment was "only" a moral obligation, the main difference in this vignette is that the movie company has an economic interest in being concerned about the quality of its movies, as it is selling them. So there is no

[9] Mean=40.23 SD=15.62.

prima facie reason why the negative outcome should be more likely than the positive one; rather, quite the contrary.

The reorganization story describes a company that has to make some organizational changes. Although those measures promise a prosperous future in the long run, they will also have short-term effects on the profits of a certain region, either positive or negative. This story was one of the first examples showing that the Knobe effect does not only occur in morally divergent scenarios, but also with "decent" actions. As previous research has demonstrated, the negative condition of this story is not assumed to be morally bad, nor is the chairman considered blameworthy for his decision (Knobe and Mendlow 2004). Therefore, an explanation of the Knobe effect should not solely focus on morality. We decided to use the reorganization story to explore whether the asymmetry in probability judgments is as pervasive as the original Knobe effect concerning intentionality ratings. Accommodated to an Austrian background, it read as follows:

> The project manager, who was responsible for developing a more simple structure for her company, reported to the chairman: "We have developed a couple of measures in order to make the company structure more compact and transparent. In the long run, this will contribute to a general increase of profits. Additionally, we expect great gainings in Styria during the next quarter [However, we expect a decrease of profits in Styria during the next quarter]." The director replies: "I don't mind Styria's balance during the next quarter at all. The main thing is that the company structures will be renewed. We are going to implement the developed measures." The reorganization runs smoothly and in Styria profits increase [decrease] during the following months.

As in Experiment 1, the participants were instructed to read each story and specify their level of agreement to statements regarding the chairman's knowledge and the chairman's intentionality using a visual analog scale ranging from "complete disagreement" to "complete agreement." After a few filler questions, they were asked to rate the probability of the occurrence of the side-effect at the time the chairman decided to implement the new program (the scale labels were "very unlikely" and "very likely"). This question resembles the specific probability question from Experiment 1.

5.3 Results

As a first step, we looked for order effects concerning the ratings of intentionality, knowledge, and probability. There were no significant results; therefore, we combined the two versions of the questionnaire with varying orders.

We found the characteristic asymmetry between the negative and the positive condition in both stories for intentionality, knowledge, and probability judgments. In other words, there were significant effects for valence for all of these three dependent variables. By way of a better illustration, these effects are summarized in Table 1 and displayed in Figs. 4 and 5.

Furthermore, we calculated two-way (valence x gender) ANOVAs to look for gender effects. In the reorganization story, there was a significant main effect of gender

Table 1 Means, standard deviation, and *t*-test statistics for intentionality, knowledge, and probability ratings in both stories

IV	Mean (SD) positive cond.	Mean (SD) negative cond.	*t*	*df*	*p*	*d*
Intentionality movie	22.13 (28.43)	78.1 (28.97)	17.97	338.00	0.000	1.96
Intentionality reorganization	24.83 (29.56)	68.87 (30.54)	13.51	338.00	0.000	1.47
Knowledge movie	78.52 (27.24)	94.08 (15.17)	6.57	280.01	0.000	0.79
Knowledge reorganization	74.29 (27.26)	90.67 (14.56)	6.99	273.40	0.000	0.85
Probability movie	83.89 (19.49)	89.42 (17.61)	2.75	337.88	0.006	0.30
Probability reorganization	79.99 (19.59)	85.06 (16.59)	2.56	336.83	0.011	0.28

concerning intentionality ratings and a significant interaction effect concerning probability ratings. In this respect, men gave higher intentionality ratings (M=50.36, SD=37.19) than women (M=42.58, SD=36.97) and showed a bigger asymmetry in probability ratings (M_{pos}=76.07, $SD_{pos:}$=21.7; M_{neg}=87.24, SD_{neg}=16.09) than women (M_{pos}=82.18, $SD_{pos:}$=17.32; M_{neg}=84.9, SD_{neg}=16.93). All of the other effects were insignificant.[10]

The correlations between knowledge and probability judgments are higher than in Experiment 1, with r=0.54 (p<0.001) in the movie story and r=0.56 (p<0.001) in the reorganization story. The mediator analyses show that the perceived probability of the side-effect could function as a partial mediator of the ESEE, as there is a significant indirect effect. The difference in knowledge ratings between the positive and the negative conditions remains significant when probability perception is statistically controlled for (direct effect).[11]

5.4 Discussion

Experiment 2 had two aims: to check whether the probabilistic side-effect effect can be replicated with different vignettes and to explore gender-specific differences concerning the ESEE. We managed to replicate the PSEE in both stories. That strengthens the evidence that the evaluation of the probability of a side-effect is

[10] Film story:
Main effect: intentionality: $F(1,340)=1.28$, $p>0.05$; knowledge: $F=0.07$, $p>0.05$; specific probability: $F=3.24$, $p>0.05$.
Interaction: intentionality: $F(1,340)=0.81$, $p>0.05$; knowledge: $F=0.03$, $p>0.05$; specific probability: $F<0.01$, $p>0.05$.
Reorganization story:
Main effect: intentionality: $F(1,340)=0.6.91$, $p<0.01$; knowledge: $F=0.09$, $p>0.05$; specific probabi lity: $F=0.87$, $p>0.05$.
Interaction: intentionality: $F(1,340)=0.04$, $p>0.05$; knowledge: $F=0.04$, $p>0.05$; specific probability: $F=5.64$, $p=0.018$.

[11] Again, we used the Preacher and Hayes (2008) SPSS macro and calculated bias-corrected bootstrap confidence intervals based on 5000 bootstrap samples.
Film story: total effect: $t(339)=6.43$, $p<0.001$, direct effect: $t(338)=5.75$, $p<0.05$ indirect effect: 99 % C.I.: 0.87 – 8.82.
Reorganization story: total effect: $t(339)=3.17$ $p<0.01$ direct effect: $t(338)=2.49$, $p<0.05$ indirect effect: 99 % C.I.=0.09 – 5.77.

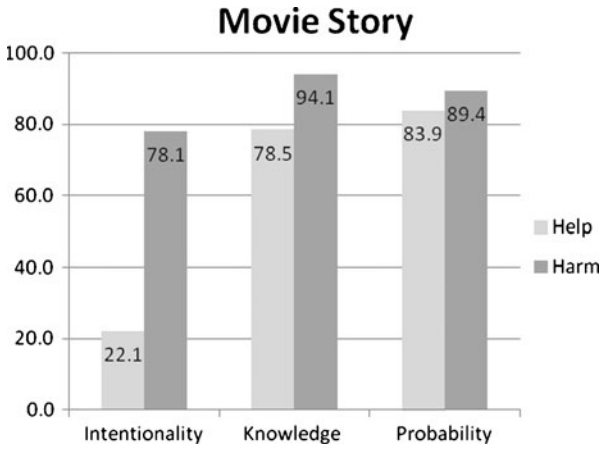

Fig. 4 Responses in all conditions of the movie story

influenced by the same sources that are characteristic for the Knobe effect. At the same time, these results also prove that the PSEE is not due to some specific attitude about harming or helping the environment.

Notably, Experiment 2 depicts that the PSEE cannot be explained by moral differences alone. This is nothing too puzzling, as it also holds true for both the original side-effect effect and the ESEE.

With regard to gender-specific characteristics of the side-effect effects, it is important to report that there are no effects concerning knowledge ratings. This supports the result from Experiment 1 that the ESEE occurs equally with women and men. Intentionality was somehow judged higher by men in the reorganization story, but that had no influence on the side-effect effect. Surprisingly, we found a significant interaction in the reorganization story for probability judgments, which shows that in this story the PSEE is stronger among men than among women. This result is not easy to understand, especially as together with the results from Experiment 1, out of four stories there is just one that leads to a bigger PSEE among men.

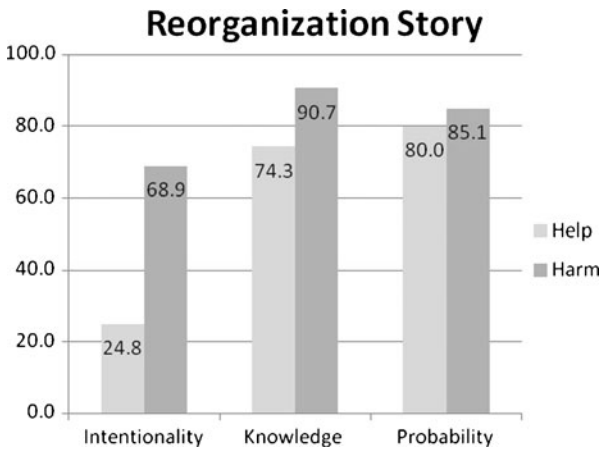

Fig. 5 Responses in all conditions of the reorganization story

Additional empirical support will certainly be required to decide about this effect's originality.

Finally, Experiment 2 also supports the results drawn from Experiment 1, showing that a difference in the perception of the probability of the side-effects is not the source of the ESEE. Although the mediation analyses suggest that probability could function as a mediator, there is still a significant direct effect of valence on knowledge ascription.

6 General Discussion

When the first paper on the Knobe effect was published nearly 10 years ago, hardly anyone would have guessed the vivid and fruitful discussion that would follow. Since then, no explanation could be given that was sufficient to consolidate all the results displayed within the existing literature. Quite the contrary: the effect that needs to be explained is no longer only about intentionality but has spread to many other concepts. Yet, it does not seem as if the end of this process has been reached. What could already be demonstrated for "decide," "want," "desire," "know," and many other variations seems to hold true for probabilities as well. There is a significant difference between the evaluation of a positive and a negative side-effect in all of these cases.

Nevertheless, it is important to note that the original effect on intentionality exceeds the effects on knowledge and probability. In both experiments, the ESEE and the PSEE are considerably smaller than the original Knobe effect. Compared to the existing literature from Beebe and Jensen (2012), our data account for a much smaller asymmetric response pattern concerning knowledge ascriptions. Regarding the ESEE, Beebe and Buckwalter (2010) reported a mean of 0.91 in the help condition and 2.25 in the harm condition, using a scale ranging from −3 to 3. Beebe and Jensen used a scale from 1 to 7 and obtained means of 3.35 and 6.37. These results equate to a difference of 43 % (d=1.99) of the scale used for Beebe and Jensen's experiment and 19 % (d=0.74) for Beebe and Buckwalter's data respectively. In comparison, we used a visual analog scale of between 1 and 100 and obtained means of 87.05 and 96.07, which corresponds to a difference of 9 % (d=0.47). This means that the ESEE reported by Beebe and Jensen was more than four times greater than the effect we observed. Interestingly, the extent of the asymmetry reported by Beebe and Jensen declines abruptly when the authors report the results of other vignettes. They only found a difference of 8 % (d=0.29) in the movie story and 10 % (d=0.39) in the reorganization story, whereas we found a difference of 16 % (d_{movie}=0.79 and $d_{reorganization}$=0.85) in both stories. (Likewise, Beebe and Jensen's findings are unique, as they report a mean beneath the midpoint of the scale used.) In this respect, the effect size found by Beebe and Buckwalter (2010) seems the most realistic and matches the effect size that we found in Experiment 2. Therefore, we conclude that the ESEE is not smaller when explored with German material.

Considering all the existing results, it appears safe to argue that the ESEE is considerably smaller than the SEE. In this respect, we found a difference of 57.83 % (d=2.33), which agrees with the results of previous studies (cf. Lanteri 2012).

Due to the small effect size of the ESEE, it seems even more interesting to study the influence of perceived probabilities on knowledge ascription. There are quite substantive correlations between knowledge ascription and probability judgments with medium to big effect sizes in Experiments 1 and 2. Moreover, the mediation analyses constitute strong evidence in favor of such an impact, as significant indirect effects can be found in all four stories. Additionally, there were also significant direct effects, which means that although the probability of the side-effect does influence the participants' knowledge ratings, it is not the source of the ESEE. It is important to consider that probability ratings are influenced in the same way as knowledge ratings. This fact could be responsible for the indirect effect that we found in Experiments 1 and 2.

As this study presents data for the SEE as well as the ESEE, this provides the opportunity to clarify whether any gender differences exist. This question gains importance within the debate on general gender differences regarding the interpretation of thought experiments (cf. in particular Stich and Buckwalter 2011; Buckwalter and Stich 2013). Inter alia, the authors report a significant gender effect concerning the ESEE. This assertion is based on a reanalysis of the data from Beebe and Buckwalter (2010), suggesting that the ESEE is stronger for female participants. In our survey it turned out that neither the SEE nor the ESEE is accompanied by a gender effect. This concurs with the existing literature insofar as there are no reports about gender effects concerning the SEE up to now, and it is highly implausible that no one has yet investigated any gender-specific differences. Therefore, our study provokes doubts about the claim that there might be gender differences concerning the ESEE. In our opinion, the result of a single study should not be overrated, considering the abundance of literature that does not mention anything about gender effects, since it would be highly unorthodox to report the fact that no effect could be detected. This phenomenon is well known as the "file drawer effect."

Interestingly, we found a gender effect concerning the PSEE in one of the four stories we used in Experiments 1 and 2. Of course, this is not enough evidence to conclude that there are overall gender-specific differences in probability perception. Rather, this result shows how easily gender effects occur when material is used that can be interpreted in many different ways.

7 Conclusion

The main purpose of this paper was to study one of the most surprising generalizations of the Knobe effect, the epistemic side-effect effect. On the one hand, we wanted to replicate both effects with German-speaking participants and examine whether there are differences between the sizes of these effects. In addition, we aimed to use our data to clarify whether gender effects occur in this subfield of experimental philosophy. On the other hand, we wanted to explore whether the ESEE is caused by different probability ratings concerning the side-effects described in the stories used.

Our results show that these effects appear with German material and that the asymmetry is considerably smaller for knowledge ratings than for intentionality judgments. Furthermore, we could not find any gender effects, which leads us to

hypothesize that the results reported by Buckwalter and Stich (2013) are probably merely caused by chance. The perceived probability mediates the willingness to ascribe knowledge to a certain extent, but this cannot explain the ESEE.

Surprisingly, we found that the evaluation of the probability of the side-effects is itself influenced by the same factors as the Knobe effect and the ESEE. Negative side-effects are judged to be more likely than positive ones. This holds true for a couple of different side-effects and also when objective probabilities would suggest otherwise. Because of the pervasiveness of significant asymmetries in probability judgments, we conclude that this represents a new generalization of the Knobe effect – a probabilistic side-effect effect (PSEE).

Further research will be required to find out whether the correlation of the perception of probability and the ascription of knowledge is due to a common cause of all side-effect effects or to the fact that one influences the other (or both to a certain extent). Furthermore, there needs to be an exploration of why there is a shift in the perception of probability depending on whether a general or a more specific question is asked. Finally, the quest remains for a satisfying explanation of the side-effect effects.

References

Adams, F., and A. Steadman. 2004a. Intentional action in ordinary language: Core concept or pragmatic understanding? *Analysis* 64(2): 173–181.

Adams, F., and A. Steadman. 2004b. Intentional action and moral considerations: Still pragmatic. *Analysis* 64(3): 268–276.

Beebe, J.R., and W. Buckwalter. 2010. The epistemic side-effect effect. *Mind and Language* 25(4): 474–498.

Beebe, J.R., and M. Jensen. 2012. Surprising connections between knowledge and action: The robustness of the epistemic side-effect effect. *Philosophical Psychology* 25(5): 689–715.

Buckwalter, W., and S. Stich. 2013. Gender and philosophical intuition. In *Experimental philosophy, vol. 2*, ed. J. Knobe and S. Nichols. Oxford: Oxford University Press.

Cullen, S. 2010. Survey-driven romanticism. *Review of Philosophy and Psychology* 1(2): 275–296.

Cushman, F., and A. Mele. 2008. Intentional action: Two-and-a-half folk concepts? In *Experimental philosophy*, ed. J. Knobe and S. Nichols, 171–188. Oxford: Oxford University Press.

Deutsch, M. 2009. Experimental philosophy and the theory of reference. *Mind and Language* 24(4): 445–466.

Egré, P. 2010. Qualitative judgments, quantitative judgments, and norm-sensitivity. [Commentary on J. Knobe's "Person as Scientist, Person as Moralist"]. *Brain and Behavioral Sciences* 33(4): 335–336.

Egré, P. 2013. Intentional action and the semantics of gradable expressions (On the Knobe Effect). In *Causation in grammatical structures*, ed. B. Copley and F. Martin. Oxford: Oxford University Press.

Feltz, A. 2007. The Knobe effect: A brief overview. *The Journal of Mind and Behavior* 28: 265–277.

Guglielmo, S., and B.F. Malle. 2010. Can unintended side effects be intentional? resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin* 36(12): 1635–1647.

Hayes, A.F. 2009. Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs* 76: 408–420.

Knobe, J. 2003. Intentional action and side-effects in ordinary language. *Analysis* 63(3): 190–193.

Knobe, J. 2004a. Folk psychology and folk morality: Response to critics. *Journal of Theoretical and Philosophical Psychology* 24(2): 270–279.

Knobe, J. 2004b. Intention, intentional action and moral considerations. *Analysis* 64(2): 181–187.

Knobe, J., and G. Mendlow. 2004. The good, the bad, and the blameworthy: Understanding the role of evaluative reasoning in folk psychology. *Journal of Theoretical and Philosophical Psychology* 24(2): 252–258.

Knobe, J., and A. Burra. 2006. The folk concepts of intention and intentional action: A cross-cultural study. *Journal of Cognition and Culture* 6(1–2):113–132.

Knobe, J., W. Buckwalter, P. Robbins, H. Sarkissian, T. Sommers, and S. Nichols. 2012. Experimental philosophy. *Annual Review of Psychology* 63(50): 72–73.
Lane, J.D., H.M. Wellman, S.L. Olson, J. LaBounty, and D.C.R. Kerr. 2010. Theory of mind and emotion understanding predict moral development in early childhood. *British Journal of Developmental Psychology* 28: 871–889.
Lanteri, A. 2012. Three-and-a-half folk concepts of intentional action. *Philosophical Studies* 158(1): 17–30.
Leslie, A., J. Knobe, and A. Cohen. 2006. Acting intentionally and the side-effect effect: "Theory of mind" and moral judgment. *Psychological Science* 17: 421–427.
Machery, E., R. Mallon, S. Nichols, and S. Stich. 2004. Semantics, cross-cultural style. *Cognition* 92(3): B1–B12.
Machery, E., R. Mallon, S. Nichols, and S. Stich. 2012. If folk intuitions vary, then what? *Philosophy and Phenomenological Research*. doi:10.1111/j.1933-1592.2011.00555.x.
Machery, E., C.Y. Olivola, and M. De Blanc. 2009. Linguistic and metalinguistic intuitions in the philosophy of language. *Analysis* 69(4): 689–694.
Mallon, R., E. Machery, S. Nichols, and S. Stich. 2009. Against arguments from reference. *Philosophy and Phenomenological Research* 79(2): 332–356.
Marti, G. 2009. Against semantic multiculturalism. *Analysis* 69(1): 42–48.
McCann, H.J. 2005. Intentional action and intending: Recent empirical studies. *Philosophical Psychology* 18(6): 737–748.
Mele, A. 1992. Recent work on intentional action. *American Philosophical Quarterly* 29(3): 199–217.
Mele, A. (ed.). 1997. *The philosophy of action*. Oxford: Oxford University Press.
Michelin, C., S. Pellizzoni, M.A. Tallandini, and M. Siegal. 2010. Evidence for the side-effect effect in young children: Influence of bilingualism and task presentation format. *European Journal of Developmental Psychology* 7(6): 641–652.
Nadelhoffer, T. 2004. On praise, side effects, and folk ascriptions of intentionality. *Journal of Theoretical and Philosophical Psychology* 24(2): 196–213.
Nagel, J. 2012. Intuitions and experiments: A defense of the case method in epistemology. *Philosophy and Phenomenological Research* 85(3): 495–527.
Nichols, S., and J. Ulatowski. 2007. Intuitions and individual differences: The Knobe effect revisited. *Mind and Language* 22(4): 346–365.
Nichols, S., S. Stich, and J.M. Weinberg. 2003. Metaskepticism: Meditations in ethnoepistemology. In *The skeptics*, ed. S. Luper, 227–247. Aldershot (Hampshire): Ashgate.
Pellizzoni, S., M. Siegal, and L. Surian. 2009. Foreknowledge, caring, and the side-effect effect in young children. *Developmental Psychology* 45(1): 289–295.
Pettit, D., and J. Knobe. 2009. The pervasive impact of moral judgment. *Mind and Language* 24(5): 586–604.
Pighin, S., J.-F. Bonnefon, and L. Savadori. 2011. Overcoming number numbness in prenatal risk communication. *Prenatal Diagnosis* 31(8): 809–813.
Preacher, K.J., and A.F. Hayes. 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods* 40: 879–891.
Sripada, C.S. 2010. The deep self model and asymmetries in folk judgments about intentional action. *Philosophical Studies* 15(2): 159–176.
Sripada, C.S., and S. Konrath. 2011. Telling more than we can know about intentional action. *Mind and Language* 26(3): 353–380.
Stich, S., and W. Buckwalter. 2011. Gender and the philosophy club. *The Philosopher's Magazine* 52: 60–65.
Sytsma, J., and J. Livengood. 2011. A new perspective concerning experiments on semantic intuitions. *Australasian Journal of Philosophy* 89(2): 315–332.
Turner, J. 2004. Folk intuitions, asymmetry, and intentional side effects. *Journal of Theoretical and Philosophical Psychology* 24(2): 214–219.
Uttich, K., and T. Lombrozo. 2010. Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition* 116: 87–100.
Weinberg, J.M., S. Nichols, and S. Stich. 2001. Normativity and epistemic intuitions. *Philosophical Topics* 29(1–2): 429–460.
Welch, B.L. 1947. The generalization of "student's" problem when several different population variances are involved. *Biometrika* 34(1–2): 28–35.
Young, L., F. Cushman, R. Adolphs, D. Tranel, and M. Hauser. 2006. Does emotion mediate the effect of an action's moral status on its intentional status? Neuropsychological evidence. *Journal of Cognition and Culture* 6: 265–78.