Organon F 26 (2) 2019: 228–248
https://doi.org/10.31577/orgf.2019.26203

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

The Significance of the Relationship between Main Effects and Side Effects for Understanding the Knobe Effect

Andrzej Waleszczyński*[a] – Michał Obidziński*[b] – Julia Rejewska*[c]

Received: 26 March 2019 / Accepted: 19 October 2019

Abstract: The characteristic asymmetry in ascribing intentionality, known as the Knobe effect, is widely thought to result from the moral evaluation of the side effect. Existing research has focused mostly on elucidating the ordinary meaning of the notion of intentionality, while less effort has been devoted to the moral conditions associated with the analyzed scenarios. The current analysis of the moral properties of the main and side effects, as well as of the moral evaluations of the relationship between them, sheds new light on the influence of moral considerations on the attribution of intentionality in the Knobe effect. The moral evaluation of the relationship between the main and side effects is significant in that under certain circumstances it cancels asymmetry in intentionality ascription.

Keywords: Asymmetry; intentional action; intentionality; Knobe effect; moral evaluation; moral properties; side effect.

* Cardinal Stefan Wyszyński University

[a] ✎ Corresponding author. Institute of Philosophy, Faculty of Christian Philosophy, Cardinal Stefan Wyszyński University, Wóycickiego 1/3, 01-938 Warsaw, Poland
✉ a.waleszczynski@uksw.edu.pl

[b] ✉ m.m.obidzinski@gmail.com

[c] ✉ julia.rejewska@gmail.com

1. Introduction

In 2003, Joshua Knobe conducted an experiment on the tendency to ascribe intentionality to actions. Currently, it is referred to in literature as the Knobe effect, according to which people have a tendency to ascribe intentionality in cases of negative, but not positive, side effects (Knobe 2003a). An understanding of the nature of moral discernment and its characteristics plays an essential role in elucidating the influence of moral considerations on the ordinary concept of intentionality. In recent years, this issue has drawn much interest, with numerous empirical studies striving to explain the observed effect (Knobe 2003a, 2003b, 2006; Nadelhoffer 2004a, 2004b, 2006; Wright and Bengson 2009; Holton 2010; Sripada 2010, 2012; Sripada and Konrath 2011; Hindriks, Douven, and Singmann 2016).

It appears that scenarios patterned after Knobe's structure of stories contain other components subject to moral evaluation in addition to the side effect, i.e. the moral value of the main effect or the moral value of the relationship between the main effect and the side effect. Furthermore, whereas the explanations of the Knobe effect offered to date have predominantly focused on the moral evaluation of the side effect, these other components may carry differential moral properties in different scenarios. Given the above theoretical premises, the central objective of the present paper is to examine the contribution of other moral considerations, such as moral evaluations of the main effect or of the relationship between the main and side effects, to ascribing intentionality in side-effect cases. Of interest here, is whether evaluations of other scenario components significantly affect the aforementioned asymmetry in ascribing intentionality. For this reason, in this paper, we are interested in whether the moral evaluations of main effect and side effect—and the relationship between them—significantly influence the attribution of intentionality to actions. We do not intend to try to explain the Knobe effect, but to examine how the moral evaluation of effects impacts the ascription of intentionality to the side effect.

2. Intentional action and the Knobe effect

In his widely discussed paper "Intentional Action and Side Effects in Ordinary Language" Knobe (2003a) Knobe presented an interesting experiment concerning ordinary intuitions associated with ascribing intentionality. He presented respondents with two scenarios which were structurally identical in terms of intentional behavior theory, the only difference being the moral value of the side effects of the agent's actions, which had not been taken into account in standard approaches. One scenario represented a "help" version with the side effect being positive, while the other one contained a "harm" version, with the side effect being negative.

The scenario with the "harm" version was as follows:

> The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.' The chairman of the board answered 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was harmed. (Knobe 2003a, 191)

And the one with the "help" version was as follows:

> The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also help the environment.' The chairman of the board answered 'I don't care at all about 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 helped. (Knobe 2003a, 191)

The respondents were asked whether the chairman of the board intentionally harmed or helped the environment (depending on the version of the scenario). It was found that they were more likely to ascribe intentionality when the side effect was negative (82%) versus positive (23%). Since then,

numerous studies and analyses have corroborated the observed asymmetry in ascribing intentionality, which has come to be known as the "Knobe effect" or the "side-effect effect." Knobe's results have been also replicated in other languages, e.g., Hindi (Knobe and Burra 2006), German (Dalbauer and Hergovich 2013), and Polish (Kuś and Maćkiewicz 2016; Waleszczyński, Obidziński and Rejewska 2018), which indicates that the Knobe effect is culture- and language-independent, and as such may be successfully studied in the Polish language.

Knobe's research was focused on the issue of intentionality. According to the *Simple View* of intentional action (SV) (Adams 1986; McCann 1987), if the agent does not intend to cause a certain effect, then she cannot bring it about intentionally. Following this line of thinking, it would be erroneous to ascribe intentionality to the side effects described in either scenarios, in which the chairman of the board makes an uncoerced decision to implement a new corporate program designed to cause a positive effect A. Thus, achieving A is clearly the chairman's objective. At the same time, the chairman has been informed (he predicts) that the initiation of the new program will also result in an additional side effect B. The chairman states that he is solely interested in achieving A and is completely indifferent to B. In other words, the chairman indicates that B is not his intention. In the two scenarios, the variable is the moral value of B, which gives rise to asymmetry in ascribing intentionality to actions leading to B. Within the SV framework, the asymmetry would be explained as erroneous attributions in the "harm" scenario. However, the situation is more complex. If the moral evaluation of the side effect is taken to influence intentionality ascriptions, it must be recognized that moral evaluation is equally applicable to the main effect. Under the circumstances, the moral evaluation of the side effect may be affected by that of the main effect, and the resulting relationship between the moral evaluations of the two effects may bear on perceptions of the side effect. It should also be borne in mind that moral evaluation is not an ordinary instance of weighing costs and benefits (Mallon 2008). Therefore, it should be examined whether a change in the nature of the relationship between the two effects may alter the ascription of intentionality in side-effect cases.

3. Explaining asymmetry in ascribing intentionality – discussion

Existing research on asymmetry ascription in side-effect cases has focused on several major aspects. Knobe (2004, 2006) explained his findings in terms of the moral evaluation of the side effect. In his opinion, people tend to ascribe intentionality when the side effect is bad, but not when it is good. Following Hindriks, this explanation shall be called the *Moral Valence Hypothesis* (MVH).

However, it should be remembered that in his seminal experiment, Knobe (2003a) formulated two questions for each scenario. One concerned the chairman's intention to cause the side effect, while the other one asked respondents how much blame (in the "harm" version) or praise (in the "help" version) the chairman deserved for bringing it about. Knobe's results showed a correlation between attributing blame and intentionality. Analysis of these results has revealed yet another asymmetry, termed the *Praise–Blame Asymmetry* (Hindriks 2008, 630), which has given rise to a new approach to the Knobe effect. Blame attribution in this context has been explored in great detail by Hindriks et al. (2016), who have reported that intentionality ascriptions depend not so much on the *Praise–Blame Asymmetry*, as on the degree of attributed blame. However, this explanation does not hold in light of Knobe and Mendlow's study (Knobe and Mendlow 2004) which has revealed an asymmetry in ascribing intentionality in the absence of a tendency to attribute blame. Moreover, there also exist situations in which intentionality is not ascribed even though the side effect is negative (bad) and blame has been apportioned (Mele 2001; Nadelhoffer 2004). Indeed, such situations have refocused the researchers' attention on the concept of responsibility (Wright and Bengson 2009; Hindriks 2011). The observed correlation between ascribing intentionality and responsibility seems to shed more light on the Knobe effect than explanations based on the concept of blame as the former makes it possible to interpret situations in which responsibility is attributed in the absence of placing blame (Knobe and Mendlow 2004; Wright and Bengson 2009).

The above explanations of the Knobe effect and their underlying hypotheses refer to other concepts (blame, responsibility) and the related

moral evaluations. A more autonomous explanation of the observed asymmetry is offered by Richard Holton (2010), who draws on the idea of a norm, and in particular norm violation. According to him, the asymmetry identified by Knobe results from the fact that people violate norms intentionally, while conforming to norms does not presuppose intentionality. In the "harm" version, the norm is violated, and consequently intentionality is ascribed, but in the "help" version the norm is observed, which is naturally interpreted as an instance of non-intentional behavior. From a philosophical perspective, the idea of a norm is similarly employed by Katarzyna Paprzycka (2014, 2015), who combines an orthodox theory of intentional action with a normative account of intentional omission. According to Paprzycka, the "harm" scenario entails an intentional omission to follow a norm. Therefore, the chairman's stated intention to achieve only the main effect does not prevent an ascription of intentional omission to observe a norm—the prerequisite for such an ascription is knowledge of the norm rather than an intention to violate it. At the same time, Paprzycka (2016) aptly observes that Holton's hypothesis about intentional norm violation (presupposing intention), presupposes intentional omission of a norm (presupposing knowledge). The main difficulty is that it is not known to what norm (if any) the respondents refer. If one assumed, as e.g., Shaun Nichols and Joseph Ulatowski (2007), that the tendency to asymmetrically ascribe intentionality in side-effect cases forms a stable pattern, the problem would only be exacerbated. This would imply that if one altered the content, but not the structure, of the scenario, then the violated norm would change as well. That would in turn mean that the respondents, in a predictable manner, each time refer to a violated norm which is different in each scenario. In other words, one would have to assume that in all experiments using Knobe's scenario structure the attitude of the respondents to the violated norm is predictable.

Finally, in F. Hindriks's *Normative Reason Hypothesis* (Hindriks 2008, 2011, 2014; Hindriks et al. 2016), the Knobe effect is explained by the agent's gradable indifference towards the side effect he has caused. According to Hindriks, in Knobe's scenario respondents perceive a certain obligation of the chairman to care about the consequences of his actions. In other words, Hindriks suggests that the chairman ignores a valid normative

reason by expressing indifference. In ordinary speech, indifference is a propositional attitude sometimes interpreted in a categorical way and sometimes in a graded way. Complete indifference would be perceived as an attitude of neutrality, with maximum caring being its polar opposite (Hindriks et al. 2016, 215–16). Hindriks treats people's assessment of the chairman's indifference as a factor affecting the degree of intentionality ascribed to him, as indicated by prior research (Mele and Cushman 2007; Phelan and Sarkissian 2008; Guglielmo and Malle 2010). The higher the chairman's indifference towards respecting a normative reason, the higher the likelihood he will be attributed blame, and thus intentionality.

4. Examining the significance of the moral evaluations of the main and side effects for the Knobe effect

The previous explanations of the asymmetry appearing in the judgments regarding the intentionality of causing the side effect were focused, on the one hand, on the moral value of this effect (Knobe 2006), violation or omission of the recognized social norm (Holton 2010; Paprzycka) or the degree of indifference of the perpetrator to the resulting side effect (Hindriks 2014, 2016) and, on the other hand, on the dependencies between judgments on intentionality and the attribution of blame or responsibility (Wright and Bengson 2009; Hindriks 2011).

Studies carried out so far seem to unify a moral property, usually bringing it to one basic element. However, the philosophical analysis of moral problems takes into account more such properties. It takes into account, for example: intention, knowledge, consequences, circumstances and voluntary actions. In the case of an action that causes the predictable side effect, for the moral evaluation of the act, the relation between the moral value of the main effect and the moral value of the side effect is also important. If the relation of the main effect to the side effect is important for the moral evaluation, it may also be important for formulating the judgments of the intentionality of causing a side effect. To this end, we have formulated a main hypothesis, which states that the moral evaluation of the effects and relations between them significantly affects the attribution of intentionality to actions.

Therefore, in order to check this hypothesis, we assume that the asymmetry of the judgments regarding the intentionality of causing a side effect, which appears in the responses to questionnaires using the structure of the story scheme proposed by Knobe, is the model. In other words, in this article we will not be interested in either the common understanding of the concept of intentional action or identifying the conditions of its application. The purpose of our research is to check the influence of moral properties on the attribution of intentionality. In our experiments, the moral property will be the relationship that occurs between the moral evaluation of the main and side effects. The disappearance of asymmetry will testify to the verification of the adopted hypothesis and the significance of the studied moral properties for the emergence of the Knobe effect.

5. Experiment 1

The first goal of the presented experiments carried out, was to answer the following question: Does the relation between main or side effect have an influence on the Knobe's effect? The research hypothesis was that a modified relationship between the moral evaluations of the main and side effects (as compared to test (N1) with the "low-value main effect and high-value side effect" condition) would affect the ascription of intentionality. The second goal, was to investigate the properties of scenarios based on Knobe's structure that change the main effect to one that is highly valued—is its effect similar to the original one?

5.1. Method

In this study, scenarios in the Polish language[1] were administered to respondents in face-to-face settings. The experiment took place at different departments in Cardinal Stefan Wyszyński University in Warsaw. Students were assigned randomly to one of three experimental conditions in the "harm" or "help" version. The experiment and questions were presented in

[1] The trouble is that there is no clear correlate of the English adverb 'intentionally' in Polish. In our experiments we used the Polish adverb, "intencjonalnie."

a traditional—paper and pencil—fashion. Trail obtained 188 participants in this experiment (32 in each versions in the second condition, and 31 in each versions in two remaining conditions) (Obidziński and Waleszczyński 2019).

The first test, (N1), with the "low-value main effect and high-value side effect" condition employed Knobe's original scenarios (Knobe 2003a), as presented in the section, *Intentional action and the Knobe effect*. Respondents presented with the "harm" and "help" scenarios were asked "Did the chairman intentionally harm the environment?" and "Did the chairman intentionally help the environment?," respectively.

The second test (N2), with the "high-value main effect and medium-value side effect" condition involved scenarios based on Knobe's structure from test N1. However, the main effect was modified so that it would be objectively highly valued. It was decided that the development of a drug for a hitherto incurable type of cancer would meet this condition. The side effect was also conceived of as a disease to align it in the same category with the main effect. At the same time, it was assumed that pneumonia as a side effect would entail a relatively low moral evaluation. According to the research hypothesis, a change in the moral evaluations of the main and side effects would shift evaluations of the relationship between these effects, which would consequently impact the ascription of intentionality in side-effect cases.

The scenario with the "harm" version was as follows:

> The vice-president of an experimental oncological hospital went to the chairman of the board and said, "We are thinking of starting the production of a new medicine. It will help us cure patients of pancreatic cancer but it will also cause pneumonia." The chairman of the board answered, "I don't care at all about causing pneumonia. I just want to cure the patients of pancreatic cancer. Let's start the production of a new medicine." They started the production of a new medicine. Sure enough, the patients came down with pneumonia.

And the one with the "help" version was as follows:

> The vice-president of an experimental oncological hospital went to the chairman of the board and said, "We are thinking of starting the production of a new medicine. It will help us cure patients

> of pancreatic cancer but it will also cure them of pneumonia." The chairman of the board answered, "I don't care at all about curing pneumonia. I just want to cure patients of pancreatic cancer. Let's start the production of a new medicine." They started the production a new medicine. Sure enough, the patients were cured of pneumonia.

The respondents were asked the question "Did the chairman intentionally cure/cause pneumonia?," depending on the scenario version. The response scale was the same as for test N1.

A subsequent test (N3), with the "high-value main effect and low-value side effect" condition was designed in order to address this interpretational difficulty. The test employed Knobe's original scenarios, but with the main and side effects reversed. In this way, both the structure of the scenarios and the moral evaluations of the two effects remained unchanged. The only modification concerned the relationship between the effects. In this experiment, the research hypothesis was that a modified relationship between the moral valuations of the main and side effects (as compared to test (N1) with the "low-value main effect and high-value side effect" condition) would affect the ascription of intentionality.

The scenario with the "harm" version was as follows:

> The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. It will help us help the environment, but it will also cause losses." The chairman of the board answered, "I don't care at all about causing losses. I just want to help the environment as much as I can. Let's start the new program." They started the new program. Sure enough, losses were caused.

And the one with the "help" version was as follows:

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

The respondents who were given the "harm" scenario were asked the question "Did the chairman intentionally cause losses?," and those who received the "help" scenario answered the question "Did the chairman intentionally increase profits?". The response scale was the same as for tests N1 and N2.

5.2. Results

First, the Shapiro–Wilk test was performed to check for normality of distribution, and it was found that none of the distributions met the normality criterion. Thus, analysis of differences between the study groups was conducted using the non-parametric Mann-Whitney *U* test. Due to the fact that all of its results were convergent with those of Student's *t*-test, the latter are presented in this paper.

The obtained data were analyzed using Student's *t*-test for independent samples to determine the presence or absence of ascription asymmetry and to establish whether the differences between the groups responding to different scenarios were statistically significant.

In the N1, the mean scores were in the "harm" version +1,36 (SD = 2.042) and in the "help" version, –1.16 (SD = 2.083) ($F^2 = 0.235$, $p = .630$; $t(60) = 4.802$, $p < .001$, and Cohen's $d_{unbiased} = 1.205$). In turn, for N2, the mean scores were in the "harm" version +0.84 (SD = 2.05) and in the "help" version –0.78 (SD = 1.879) ($F = 0.666$, $p = .406$; $t(62) = 3.306$, $p = .002$, and Cohen's $d_{unbiased} = 0{,}817$). In the N3 the mean scores were +1.65 (SD = 1.644) in the "harm" version and +0.39 (SD = 1.856) in the "help" version ($F = 1.437$, $p = .235$; $t(60) = 2.825$, $p = .006$, and Cohen's $d_{unbiased} = 0.709$).

The result of *t*-test, for the differences between "harm" and "help" scores absolute values in N1 ($M_{N1} = 4.387$, $SD_{N1} = 1.283$) and N2 ($M_{N2} = 3.687$, $SD_{N2} = 1.575$) was not significant: $F = 1.507$, $p = 0.264$; $t(61) = 1.93$, $p = 0.058$. Finally, the results of *t*-tests for differences in mean scores in "help" story judgment, between all experimental conditions were tested. For groups N1 and N3: $F = 0.709$, $p = .403$; $t(60) = –3.090$, $p = .003$, and Cohen's $d_{unbiased} = 0.775$. For groups N2 and N3: $F= 0.030$, $p = .863$; $t(61) = –2.482$, $p = .016$, and Cohen's $d_{unbiased} = 0.618$.

[2] Fisher homogeneity test.

5.3. Discussion

The obtained results support the research hypothesis that the relationship between the moral evaluations of the main and side effects has a significant influence on the symmetry of intentionality ascriptions in side-effect cases. Moreover, there was significant difference between "help" score in groups (N1) with the "low-value main effect and high-value side effect" condition and (N3) with the "high-value main effect and low-value side effect" condition. The reversal of the main and side effects cancels the attribution asymmetry reported for the original scenario versions. Taking into account the fact that the main effect/side effect relation was the only thing that differentiates the two conditions it is very possible that the observed lack of Knobe effect is due to the given experimental manipulation. Moreover, the new scenario based on the Knobe scenario turn out similar effects to the standard Knobe's scenario, thus it was contradictory to our assumption. However, the probability value for this analysis is very close to the level of significance. Moreover, taking into account our assumptions, the one-tailed test result is significant ($p = 0{,}029$).

6. Experiment 2

In the second experiment we are investigating, whether the difference observed in the first experiment will appear once again in the more random sample—thus supporting the hypothesis. Moreover, once again, modification of scenario was tested.

6.1. Method

In this study scenarios in the Polish language were administered to respondents in face-to-face settings. The participants were random people encountered in the vicinity of the Warszawa Śródmieście and Główna Railway Stations as well as the Łódź Kaliska Railway Station. Participants were assigned randomly to one of the experimental conditions in the "harm" or "help" version. The experiment and questions were presented in the traditional—paper and pencil—fashion. Trail obtained 186 participants in this experiment (31 in each versions in all three conditions) (Obidziński and

Waleszczyński 2019). The used methodology was identical to the one used in experiment 1.

6.2. Results

Once again, the Shapiro–Wilk test was performed to check for normality of distribution, and it was found that none of the distributions met the normality criterion. Thus, analysis of differences between the study groups was conducted using the non-parametric Mann-Whitney U test. Again, because of convergent results of both tests, the student's t-test will be used.

In the N1, the "harm" and "help" versions, the mean scores were +1.94 (SD = 1.731) and –1.06 (SD = 2.265), respectively, on a seven-point scale ranging from +3 (definitely yes) to –3 (definitely no), with 0 designated as "hard to say." Statistical significance was confirmed by Student's t-test ($F = 5.153$, p = .027; $t(56.130) = 5.860$; p < .001) and Cohen's $d_{unbiased}$ (1.47). Thus, as expected, the study revealed a statistically significant Knobe effect. In turn, for N2, the mean score for the "harm" version was +0.36 (SD = 2.303), and that for the "help" version was –0.39 (SD = 2.14). While the results revealed asymmetry in ascribing intentionality, it was no longer statistically significant (F = 0.699, p = .406; $t(60) = 1{,}314$; p = .194). In the N3, the mean score for the "harm" version was +0.26 (SD = 1.57), and that for the "help" version was +0.16 (SD = 2.208). Thus, the results of the two scenarios were convergent and indicative of symmetry in ascribing intentionality ($F = 7.988$, p = .006; $t(54.164) = 0.199$; p = .843).

The result of t-test, for the differences between "harm" and "help" scores absolute values in N1 ($M_{N1} = 4.613$, $SD_{N1} = 1.202$) and N2 ($M_{N2} = 3.774$, $SD_{N2} = 1.499$) was significant: F = 1.555, p = 0.232; t(61) = 2.43, p = 0.018, $d_{unbiased} = 0.61$. Finally, the results of t-tests for differences in mean scores in "help" story judgment, between all experimental conditions was tested. A significant difference was observed only for N1 and N3 conditions: F = 0.004, p = .953; t(60) = –2.158, p = .035, and Cohen's $d_{unbiased} = 0.541$.

6.3. Discussion

The obtained results support the research hypothesis that the relationship between the moral evaluations of the main and side effects has

a significant influence on the symmetry of intentionality ascriptions in side-effect cases. There were no significant differences between the "harm" and "help" versions in group (N3) with the "high-value main effect and low-value side effect" condition. Moreover, there was significant difference between "help" scores in N1 and N3 groups. The reversal of the main and side effects cancels the attribution asymmetry reported for the original scenario versions. Taking into account the fact that the main effect/side effect relation was the only thing that differentiates the two conditions it is very possible that the observed lack of Knobe effect is due to given experimental manipulation. Second, there was a significant difference between group (N1) with the "low-value main effect and high-value side effect" condition and group (N2) with the "high-value main effect and medium-value side effect" condition results. It supports our assumption that changing the main effect on one valued higher will affect the asymmetry.

7. General discussion

The point of reference for the present study was the Knobe effect, or asymmetry in ascribing intentionality in side-effect cases. The results of group (N2) with the "high-value main effect and medium-value side effect" condition indicate that intentionality ascriptions may be affected not only by the agent's indifference towards the consequences of his actions, but also by a change in the moral evaluations of the main and side effects. Test (N3) with the "high-value main effect and low-value side effect" condition has corroborated the influence of the examined moral properties on intentionality attributions and made it possible to elucidate their nature. It has been found that of greatest significance is the relationship between the moral valuations of the main and side effects. Indeed, this relationship is critical to asymmetry in intentionality ascriptions. The results of test (N3) with the "high-value main effect and low-value side effect" condition for the "help" version are significantly statistically different from those for test (N1) with the "low-value main effect and high-value side effect" condition. However, of particular importance is the fact that symmetry was obtained by a radical increase in intentionality ascriptions in the "help" version, which must be surprising from the SV perspective.

Previous efforts to explain the Knobe effect were more focused on intentionality ascriptions in the "harm" version, as those ascriptions appeared to be inconsistent with the SV. The experiments presented in this paper shed new light on prior studies exploring the notion of intentional action. The aforementioned findings from works analyzing blame apportioning seem deficient as the emergence of intentionality ascriptions in the "help" version would entail blame attribution, which is a contradiction in terms in light of the meaning of the notions of blame and morally positive effects. Therefore, the Knobe effect cannot be explained by blame apportioning, and in particular by the *Praise–Blame Asymmetry*, which only reveals an existing correlation emerging under certain specific circumstances. As regards Hindriks's *Normative Reason Hypothesis*, indifference towards the side effect should be acknowledged as a significant factor in ascribing intentionality, but it is nevertheless secondary to the relationship between the moral evaluations of the main and secondary effects. Already test (N2) with the "high-value main effect and medium-value side effect" condition showed that a change in those evaluations influenced the extent of ascribed intentionality with respect to test N1. It may be expected that variation in the degree of indifference may additionally modify intentionality attributions, but that factor is unlikely to be decisive in accounting for the observed attributional asymmetry. Indeed, it seems that Hindriks overestimated the role of indifference in explaining the Knobe effect. Also Holton's and Paprzycka's proposals do not seem to hold in light of the presented new experimental results. While their findings explain intentionality ascriptions in the "harm" version, both authors' hypotheses would be falsified if applied in the "help" version as causing a positive side effect could hardly be shown to violate any moral norms.

On the other hand, it should be noted that the presented evidence does not contradict Knobe's MVH. In explaining attribution asymmetry, Knobe proposed that it was influenced by the moral evaluation of the side effect, which is correct, but does not account for the other factors at play. While Knobe was right that moral considerations, and especially the moral evaluation of the side effect impact the ascription of intentionality in bringing it about, it has been found here that the influence of moral considerations and the moral evaluation of effects is more complex than previously thought.

The Knobe effect could be explained by a new hypothesis, proposed here as the *Wide Moral Valence Hypothesis*, according to which asymmetry in ascribing intentionality in side-effect cases is attributable to one main underlying cause, which is the moral evaluation of the relationship between the moral values associated with the main and side effects. The existence of this factor, that is, moral properties affecting the way people perceive complex situations, has been indicated by P. Egré and F. Cova (2015). They reported that the moral considerations associated with negatively valenced concepts, such as death, and positively valenced ones, such as survival, bear significantly on the way people think and perceive the world, and consequently, on the way they arrive at their evaluations. The mechanism used in Egré and Cova's study, that is, reversing the order of responses, did not affect the Knobe effect (Nichols and Ulatowski 2007). However, the present test (N3) with the "high-value main effect and low-value side effect" condition did produce results somewhat convergent with Egré and Cova's work in terms of altering the valence of evaluations. Analysis of Egré and Cova's findings in conjunction with the present evidence suggests that along with their positive and negative aspects, the effects of actions have additional attributes in the form of moral properties. If one rejects the hypothesis about the existence of the moral properties of effects, then in the "help" version of test (N3) with the "high-value main effect and low-value side effect" condition the relationship between the positive main effect and the positive side effect would remain identical to the analogous relationship from test (N1) with the "low-value main effect and high-value side effect" condition in terms of moral evaluation. If the ascription of intentionality were influenced solely by the positive dimension of effects, then the reversal (swapping) of the main and side effects should not significantly affect the respondents' ascriptions of intentionality to the agent causing the side effect. However, such a reversal did in fact have a significant impact on intentionality attributions. This means that the positive dimension of effects must also have some moral properties. Given that an analogous relationship exists between the positive and negative effects, it may be argued that the emergence of the Knobe effect depends on the relationship between the moral properties of the main and side effects.

Two questions remain open. One concerns the way in which the moral properties of effects are discerned, and the other one the way in which their significance for a given situation is determined, or "measured." It should be remembered that such discernment or "measurement" do not have to be made in a purely rational way and that they do not amount to a simple weighing of costs and benefits (Machery 2008; Mallon 2008). Therefore, it cannot be excluded that in situations where negative moral values are discerned, the process of making ordinary moral evaluations is governed by mechanisms that in some ways differ from those governing evaluations of situations characterized by positive moral values. These issues certainly require further study.

8. Summary

The objective of the present study was not so much to provide another explanation for the Knobe effect, as to test the hypothesis according to which moral evaluations of the main and side effects and the relationship between them significantly influence the attribution of intentionality to actions. Furthermore, acknowledging the crucial role of such moral evaluations, it seems reasonable to propose that certain situations are characterized by specific moral properties. It has been found that in cases of side effects the ascription of intentionality (which is distinct from passing a moral judgment) depends not only on whether the effects in question are positive or negative, but also on whether they are perceived to have positive or negative moral value. Of greatest importance is the relationship between the moral evaluations of the main and side effects. The underlying cause of this finding may be theoretically determined and identified as a factor that is crucial to human intuitions and judgments. This implies that the ordinary meaning of words is associated with certain moral properties that underpin moral evaluations. As it was shown in test (N3) with the "high-value main effect and low-value side effect" condition, those properties play a significant role in intuitions of non-ethical nature. The identification of such properties, which requires further study, constitutes a challenge to moral language and metaethical theories. It may well be that moral language goes beyond utterances containing terms such as duty, obligation,

and responsibility, and categories such as good/bad and praise/blame, and that it encompasses a wide spectrum of utterances and words exhibiting certain moral properties. Using the language of psychology, one could argue for the existence of a mechanism of axiological attribution which would associate different states of affairs (situations, normative systems), actions, or even individuals, with specific moral properties, which may be additionally modified by other moral properties inherent in ordinary utterances.

In conclusion, it should be added that the results of the present experiments suggest that the Knobe effect is mostly attributable to the moral evaluation of the relationship between the moral properties of the main and side effects. Previous replications of Knobe's seminal experiment were successful because they held the moral evaluation of the relationship between the two effects constant, and so the ascription asymmetry was reproduced. However, the current experiments, and in particular test (N3) with the "high-value main effect and low-value side effect" condition, showed that a modification of the moral evaluation of the relationship between the main and side effects may cancel the Knobe effect.

References

Adams, Frederick. 1986. "Intention and Intentional Action: The Simple View." *Mind and Language* 1 (4): 281–301. https://doi.org/10.1111/j.1468-0017.1986.tb00327.x

Dalbauer, Nikolaus, and Andreas Hergovich. 2013. "Is What is Worse More Likely? The Probabilistic Explanation of the Epistemic Side-Effect Effect." *Review of Philosophy and Psychology* 4 (4): 639–57. https://doi.org/10.1007/s13164-013-0156-1

Egré, Paul, and Florian Cova. 2015. "Moral Asymmetries and the Semantics of 'Many.'" *Semantics and Pragmatics* 8 (13): 1–45. https://doi.org/10.3765/sp.8.13

Guglielmo, Steve, and Bertram F. Malle. 2010. "Can Unintended Side Effects be Intentional? Resolving a Controversy over Intentionality and Morality." *Personality and Social Psychology Bulletin* 36 (12): 1635–47. https://doi.org/10.1177/0146167210386733

Hindriks, Frank. 2008. "Intentional Action and the Praise Blame Asymmetry." *Philosophical Quarterly* 58 (233): 630–41. https://doi.org/10.1111/j.1467-9213.2007.551.x

Hindriks, Frank. 2011. "Control, Intentional Action, and Moral Responsibility." *Philosophical Psychology* 24 (6): 787–801. https://doi.org/10.1080/09515089.2011.562647

Hindriks, Frank. 2014. "Normativity in Action: How to Explain the Knobe Effect and Its Relatives." *Mind and Language* 29 (1): 51–72. https://doi.org/10.1111/mila.12041

Hindriks, Frank, Igor Douven, and Henrik Singmann. 2016. "A New Angle on the Knobe Effect: Intentionality Correlates with Blame, not with Praise." *Mind and Language* 31 (2): 204–20. https://doi.org/10.1111/mila.12101

Holton, Richard. 2010. "Norms and the Knobe Effect." *Analysis* 70 (3): 417–24. https://doi.org/10.1093/analys/anq037

Knobe, Joshua. 2003a. "Intentional Action and Side Effects in Ordinary Language." *Analysis* 63 (3): 190–94. https://doi.org/10.1093/analys/63.3.190

Knobe, Joshua. 2003b. "Intentional Action in Folk Psychology: An Experimental Investigation." *Philosophical Psychology* 16 (2): 309–24. https://doi.org/10.1080/09515080307771

Knobe, Joshua. 2004. "Folk Psychology and Folk Morality: Response to Critics." *Journal of Theoretical and Philosophical Psychology* 24 (2): 270–79. https://doi.org/10.1037/h0091248

Knobe, Joshua. 2006. "The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology." *Philosophical Studies* 130 (2): 203–31. https://doi.org/10.1007/s11098-004-4510-0

Knobe, Joshua, and Arudra Burra. 2006. "The Folk Concepts of Intention and Intentional Action: A Cross-Cultural Study." *Journal of Cognition and Culture* 6 (1–2): 113–32. https://doi.org/10.1163/156853706776931222

Knobe, Joshua, and Gabriel 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): 252–58. https://doi.org/10.1037/h0091246

Kuś, Katarzyna, and Bartosz Maćkiewicz. 2016. "Z rozmysłem, ale nie specjalnie. O językowej wrażliwości filozofii eksperymentalnej." *Filozofia Nauki* 24 (3): 73–102.

Machery, Edouard. 2008. "The Folk Concept of Intentional Action: Philosophical and Experimental Issues." *Mind and Language* 23 (2): 165–89. https://doi.org/10.1111/j.1468-0017.2007.00336.x

Mallon, Ron. 2008. "Knobe versus Machery: Testing the Trade-off Hypothesis." *Mind and Language* 23(2): 247–55. https://doi.org/10.1111/j.1468-0017.2007.00339.x

McCann, Hugh J. 1987. "Rationality and the Range of Intention." *Midwest Studies in Philosophy* 10 (1): 191–211. https://doi.org/10.1111/j.1475-4975.1987.tb00540.x

Mele, Alfred R. 2001. "Acting Intentionally: Probing Folk Notions." In *Intentions and Intentionality: Foundations of Social Cognition*, edited by Bertram Malle, L. J. Moses, and Dare Baldwin, 27–43. Cambridge: MIT Press.

Mele, Alfred R., and Fiery Cushman. 2007. "Intentional Action, Folk Judgments, and Stories: Sorting Things out." *Midwest Studies in Philosophy* 31 (1): 184–201. https://doi.org/10.1111/j.1475-4975.2007.00147.x

Nadelhoffer, Thomas. 2004a. "Blame, Badness, and Intentional Action: A Reply to Knobe and Mendlow." *Journal of Theoretical and Philosophical Psychology* 24 (2): 259–69. https://doi.org/10.1037/h0091247

Nadelhoffer, Thomas. 2004b. "On Praise, Side Effects, and Folk Ascriptions of Intentionality." *Journal of Theoretical and Philosophical Psychology* 24 (2): 196–213. https://doi.org/10.1037/h0091241

Nadelhoffer, Thomas. 2004c. "The Butler Problem Revisited." *Analysis* 643 (3): 277–84. https://doi.org/10.1111/j.0003-2638.2004.00497.x

Nadelhoffer, Thomas. 2006. "Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality." *Philosophical Explorations* 9 (2): 203–19. https://doi.org/10.1080/13869790600641905

Nichols, Shaun, and Joseph Ulatowski. 2007. "Intuitions and Individual Differences: The Knobe Effect Revisited." *Mind and Language* 22 (4): 346–65. https://doi.org/10.1111/j.1468-0017.2007.00312.x

Obidziński, Michał, and Andrzej Waleszczyński. 2019. "The Significance of the Relationship between Main Effects and Side Effects for Understanding the Knobe Effect: Database." OSF. February 2. osf.io/ky3re.

Paprzycka, Katarzyna. 2014. "Rozwiązanie problemu Butlera i wyjaśnienie efektu Knobe'a." *Filozofia Nauki* 22 (2): 73–96.

Paprzycka, Katarzyna. 2015. "The Omissions Account of the Knobe Effect and the Asymmetry Challenge." *Mind and Language* 30 (5): 550–71. https://doi.org/10.1111/mila.12090

Paprzycka, Katarzyna. 2016. "Intention, Knowledge, and Disregard of Norms: The Ommisions Account and Holton's Account of the Assymetric Intentionality Attributions." In *Uncovering Facts and Values: Studies in Contemporary Epistemology and Political Philosophy*, edited by Adrian Kuźniar and Joanna Odrowąż-Sypniewska, 204–33. Leiden - Boston: Brill Rodopi. https://doi.org/10.1163/9789004312654_015

Phelan, Mark T., and Hagop Sarkissian. 2008. "The Folk Strike Back; Or, Why You Didn't Do It Intentionally, though It Was Bad and You Knew It." *Philosophical Studies* 138 (2): 291–98. https://doi.org/10.1007/s11098-006-9047-y

Sripada, Chandra Sekhar 2010. "The Deep Self Model and Asymmetries in Folk Judgments about Intentional Action." *Philosophical Studies* 151 (2): 159–76. https://doi.org/10.1007/s11098-009-9423-5

Sripada, Chandra Sekhar. 2012. "Mental State Attributions and the Side-Effect Effect." *Journal of Experimental Social Psychology* 48 (1): 232–38. https://doi.org/10.1016/j.jesp.2011.07.008

Sripada, Chandra Sekhar, and Sara Konrath. 2011. "Telling More Than We Can Know About Intentional Action." *Mind and Language* 26 (3): 353–80. https://doi.org/10.1111/j.1468-0017.2011.01421.x

Waleszczyński, Andrzej, and Michał Obidziński, and Julia Rejewska. 2018. "The Knobe Effect from the Perspective of Normative Orders." *Studia Humana* 7 (4): 9–15. https://doi.org/10.2478/sh-2018-0019

Wright, Jennifer C., and John Bengson. 2009. "Asymmetries in Folk Judgments of Responsibility and Intentional Action." *Mind and Language* 24 (1): 24–50. https://doi.org/10.1111/j.1468-0017.2008.01352.x