Studia Humana
Volume 7:4 (2018), pp. 9—15
DOI: 10.2478/sh-2018-0019

The Knobe Effect From the Perspective of Normative Orders

Andrzej Waleszczyński

Cardinal Stefan Wyszynski University,
Warsaw, Poland

e-mail: a.waleszczynski@uksw.edu.pl

Michał Obidziński

Cardinal Stefan Wyszynski University,
Warsaw, Poland

e-mail: m.m.obidzinski@gmail.com

Julia Rejewska

Cardinal Stefan Wyszynski University,
Warsaw, Poland

e-mail: julia.rejewska@gmail.com

Abstract:
The characteristic asymmetry in the attribution of intentionality in causing side effects, known as the Knobe effect, is considered to be a stable model of human cognition. This article looks at whether the way of thinking and analysing one scenario may affect the other and whether the mutual relationship between the ways in which both scenarios are analysed may affect the stability of the Knobe effect. The theoretical analyses and empirical studies performed are based on a distinction between moral and non-moral normativity possibly affecting the judgments passed in both scenarios. Therefore, an essential role in judgments about the intentionality of causing a side effect could be played by normative competences responsible for distinguishing between normative orders.
Keywords: intentional action, Knobe effect, Joshua Knobe, normativity, normative orders, normative competences.

1. Introduction

In this article we will look for an answer to the following problem: does the way of thinking about the intentionality of causing a side effect in morally negative situations affect the way of thinking about the intentionality of causing a side effect in morally positive situations, or vice versa? This question is interesting in view of the fact that the so-called Knobe effect is seen as a stable model describing human judgments about the intentionality of action [19], one of the reasons for this being that none of the numerous studies performed thus far have managed to falsify the effect. One should ask, however, what – apart from the findings of empirical studies – supports the thesis about stability of the model of intentionality attributions revealed in the Knobe effect. What theoretical arguments support this thesis?

2. The Attribution of Intentionality

Gilbert Harman [5] was one of the first scholars to discuss the difficulty related to the everyday use of the concept of intentional action. It is related to asymmetrical attribution of intentionality in causing an effect occurring in result of an accidental action. A broader discussion of this issue can be found in the works of Ronald J. Butler [3], who observed a tendency in judgments about intentionality that was difficult to explain despite the existence of analogical factors usually taken into account when such actions are analysed. In a new form, the problem resurfaced in studies performed by Joshua Knobe [10] which revealed a tendency that is now referred to in literature as the Knobe effect, or the side-effect effect.

In 2003, Knobe performed an experiment in which participants were randomly assigned a questionnaire describing one of the following scenarios:

The HARM scenario 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. [10, p. 191]

The scenario was followed by two questions:

1. Did the chairman intentionally harm the environment?
2. How much blame does the chairman deserve for what he did?

The HELP scenario 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. [10, p. 191]

The scenario was followed by two questions:

1. Did the chairman intentionally help the environment?
2. How much praise does the chairman deserve for what he did?

The study revealed that participants attributed intentionality much more readily when the side effects were negative (82%) than when they were positive (23%). Since the article was published, many comments have been made, and a number of studies have been performed in order to explain this phenomenon.

3. Attempts at Explaining the Knobe Effect

One of the standpoints which have become a permanent element in discussions around the Knobe effect is one which explains the observed asymmetries with moral factors [11]. This standpoint has

its advocates both among philosophers [14], [17] and psychologists [4], [12]. Correlations have been sought between intentionality attributions and moral judgments. A great deal of attention has been paid to the relationship between the attribution of intentionality and the attribution of guilt [13], [17], [15], [16], [18], [7], [6]. Some substantiations take into account the essential role of moral factors focused on norms and explained the attribution of intentionality with their violation [8] or intentional omission [20], [21]. Authors focusing on the role of moral arguments in explaining the observed phenomena paid less attention to subtleties related to categorisations or practical application of the concept of intentional action [2], [1], as they proved to be insufficient to explain the observed asymmetries [9], [19].

Analyses performed so far have either sought to provide an explanation which usually referred to one aspect of the issue under examination or described only some of the processes or existing correlations. It also seems that the very attitude to explaining the existing asymmetries is largely focused on subtle nuances in understanding the concept of intentional action. It is therefore interesting to use the category of prediction in order to understand the attribution of intentionality in causing side effects. In the cases of the asymmetry analysed here, it is predictions, or expectations held within the framework of a normative order embraced by the subject, that affect judgments about the intentional or non-intentional character of an action. It is worth noting that actions are based on cognitive predictions which cannot be reduced to intentions or designs [22]. Predictions are also related to the need to reduce normative tension and uncertainty. Therefore, the cause of a particular action may be seen as the need to minimize normative uncertainty [23, pp. 16-17].

According to Waleszczyński, in the search for an explanation of the asymmetry in the attribution of intentionality in causing morally positive or negative effects, it would be sufficient to point to the existence of two types of normativity: a moral and a non-moral one. This would explain most of the difficulties involved in the asymmetry discussed here. First of all, however, one should consider why any tension between the two types of normative orders should exist at all. Trying to explain the asymmetry in judgments about the intentionality of actions in the context of morally negative or positive effects, Waleszczyński has proposed the following solution [24]. With regard to the question about the intentionality of action, there are two normative orders, i.e. a moral and a non-moral one, in which different conditions apply for using the concept of intentional action. In the conditions of moral normativity, subject S_1 may be considered the originator of a good effect X_1 if effect X_1 was desired and foreseen, i.e. intended. In order to consider subject S_1 the originator of a negative effect X_2, it is enough for the particular effect X_2 to have been foreseen by subject S_1. In the conditions of moral normativity, the attribution of authorship is equivalent to intentional causation of a particular effect. It should be remembered, however, that there are various conditions for causing a morally good or bad effect within the framework of moral normativity. However, in the conditions of non-moral normativity, moral authorship (the causing of an effect which is endowed with certain moral qualities and conditions for judgment) should be distinguished from the intentionality of causing a particular effect. Therefore, in order to conclude that subject S_1 intentionally caused effect $X_{1\text{-}2}$, it is necessary to make sure whether or not he had the intention of causing effect $X_{1\text{-}2}$.

Taking the above distinctions into account, the explanation of the problem of asymmetry would be as follows: regarding the question about the intentionality of action, two normative orders overlap in which different conditions apply for using the concept of intentional action. When we are dealing with causing a good effect, the normative conditions governing the attribution of intentionality in both types of normativity coincide. In situations where the effect is morally negative, however, we may be dealing with a normative tension caused by different conditions for using the concept of intentional action, depending on the type of normativity. The distinction between two types of normativity provides a simple explanation of the asymmetry revealed in the Knobe effect. The solution proposed here relies largely on intuitions generally acknowledged in ethics.

According to Waleszczyński, however, the problem involved in the Knobe effect occurs at a certain metalevel and is related to normative competences, which enable us to distinguish between

various types of normativity. It is the normative competences which would determine according to which of the normative orders the problem is to be solved. Only after the normative order has been selected are "moral" competences or "cause-and-effect" competences employed, as applicable. The significance of moral competences would be particularly important in the case of passing judgments on the intentionality of action. When making such judgments, the conditions for applying the concept of intentional action corresponding to the two types of normative orders overlap. It is the ability to decide which type of normativity a particular question refers to and to identify the applicable conditions that would determine the judgments issued or the attribution of intentionality.

4. Discussion of the Sequence Hypothesis

If the division into two normative orders, a moral and a non-moral (cause-and-effect) one, is accepted, and considering studies on the Knobe effect performed so far, the following assumption should be made: participants who analyse the HARM condition scenario apply moral normativity, as in the case of a morally negative effect, they point to knowledge as the substantiation for the attribution of intentionality in causing that effect [24, pp. 122-4]. We do not know, however, what normative order is applied by participants who analyse the HELP condition scenario. The failure to attribute intentionality in causing a morally positive effect is substantiated by saying that the chairman did not want to or did not intend to cause such an effect. The reference to intentions behind actions and the assumptions we make in the substantiation suggests that when solving the problem, the participants could have been applying moral normativity, non-moral normativity, or both.

In order to check the above assumptions, we have decided to investigate the sequence hypothesis. The test consists in participants first being given one questionnaire, and another one after they have answered the first one. This way, we can see if the sequence in which the questionnaires are answered affects the occurrence of the Knobe effect. The sequence thesis has already been tested by Nichols and Ulatowski [19], but only to a limited extent. Their study was carried out online, and the participants could not correct their answers. The authors of the experiment did not reveal detailed results after the study was completed, but only stated that the sequence in which the questionnaires were answered did not affect the occurrence of the Knobe effect.

The matter does not seem to be as simple as this, however. If the participants prefer moral normativity when analysing the HARM condition scenario, and if we accept the principle that similar problems are solved in a similar way, the analysis of the HELP condition scenario will begin with preference for moral normativity. If this is the case, then the Knobe effect should appear in a "strong" form in both conditions, and individual judgments should be prevailingly asymmetrical. If, however, we do not know in reference to what normativity participants analyse the HELP condition scenario (there being three possibilities), then it will also be difficult to settle the preference of which normativity will come first when analysing the HARM condition scenario. If, however, the HELP condition scenario is not analysed at least by some of the participants in terms of moral normativity, then overall group results should reveal the Knobe effect in a "weaker" form, while individual results should be less asymmetrical.

Our experiment was designed as follows. The study was carried out in the form of a direct survey in which questionnaires in the Polish language were presented to passers-by encountered in the vicinity of Warszawa Główna, Warszawa Śródmieście, and Łódź Kaliska railway stations. The survey was carried out in two groups: Group 1 (HARM-HELP) and Group 2 (HELP-HARM). Each group included 31 participants. The participants were first given a questionnaire presenting the story with one condition, and after they completed it, the story with the other condition was revealed. Both stories were presented on the same page and were followed by a brief explanation on how to make corrections if a wrong answer had been given. When answering the questionnaire with the other condition, the participant could see both stories and his or her answers directly. The survey used the original Knobe stories [10], the content of which is presented in the *Attribution of*

Intentionality section. In the HARM condition questionnaire, participants had to answer one question: "Did the chairman intentionally harm the environment?"; in the HELP condition questionnaire, the question was: "Did the chairman intentionally help the environment?". Answers were given on a seven-point scale, where "+3" meant "Absolutely Yes", "-3" meant "Absolutely Not", and "0" meant "Hard to Say".

First, an analysis was performed within each group by looking at the answers of the same persons presented with the two questionnaire types (HARM and HELP). The first group began with the HARM scenario, and the other was first asked to complete the HELP scenario questionnaire. As the distribution of answers significantly differs from normal distribution, nonparametric tests were used in the analyses. The average and standard deviation for individual groups and conditions are presented in Table 1; results of the Mann-Whitney U test are presented in Table 2.

Table 1
Description of statistical results in HARM and HELP questionnaires by group

	N	M_{Harm}	SD_{Harm}	M_{Help}	SD_{Help}
Group 1 (HARM-HELP)	31	1,936	1,731	-1,387	2,108
Group 2 (HELP-HARM)	31	0,807	2,428	-1,065	2,265

Table 2
Results of the Wilcoxon test of differences between results within the same group in both questionnaire types

	Z	P	r Cohena
Group 1 (HARM-HELP)	-4,258	< 0,001	0,541
Group 2 (HELP-HARM)	-2,773	0,006	0,352

Test results of analyses using the Wilcoxon test show that in both groups the answers were asymmetrical. The effect size for Groups 1 and 2 were large and average, respectively. The difference seems to be greater in the group starting with the HARM scenario. To see if this difference is statistically significant, differences were calculated for each individual, and both groups were compared using the Mann-Whitney U test. The results are presented in the table below.

Table 3
Results of the U test comparing differences between results in the first and second questionnaire within the groups

	Z	P	r Cohena
Test U Manna-Whineya	-5,193	< 0,001	0,660

The observable difference proves to be statistically significant, and the size effect of the sequence in which the questionnaires were answered is large (which means that when the HARM scenario is analysed first, the Knobe effect is greater). Finally, to see if the differences occur in both study conditions or in only one of them, the results of each group in the HARM and HELP scenario were compared. The results of this analysis are presented in Table 4.

Table 4
Results of the U test between the groups separately for Harm and Help scenarios

	Z	P	r Cohena
Condition HARM	-1,776	0,076	0,226
Condition HELP	-0,488	0,625	-

As can be seen from the results presented above, no statistically significant differences were observed. The statistical tendency in the case of the HARM scenario suggests, however, that if a larger sample were tested, the statistical difference would probably be significant.

Final individual answers in terms of asymmetry were as follows. In Group 1 (HARM-HELP), asymmetrical answers represented 61.3%, symmetrical answers accounted for 19.35%, including four "Yes's" and two "No's,"; answers with one "0", meaning "Hard to Say", represented 19.35%. Three persons used the option to change their answer. Two persons changed their answer from an asymmetrical one to a symmetrical one, with one "0" answer. One person changed his or her answer from a symmetrical to an asymmetrical one. In Group 2 (HELP-HARM) there were 41.9% asymmetrical answers and 45.2% symmetrical answers, including five "Yes's", seven "No's", and two "0s", while answers with one "0" represented 12.9%. Just as in Group 1, the option to change the answer was used by three persons. Two persons changed their answer from a symmetrical one (including one with two "0" answers) to an answer with one "0". One person changed his or her answer from a symmetrical to an asymmetrical one.

5. Summary

An analysis of the findings suggests that in spite of the occurrence of the Knobe effect in group results, a statistical difference exists between the two groups. Individual results are interesting as well. In Group 2, symmetrical answers were more frequent than asymmetrical ones, and compared to answers in Group 1, there were twice as many. As the sample was not large enough, a more in-depth statistical analysis of this aspect was not possible.

The study we have performed and the results we have obtained suggest that the thesis about the existence of two normative orders and their impact on the attribution of intentionality in causing a side effect becomes more significant. Results in Group 2 proved to be interesting as asymmetrical answers only represented 41.9% of the total. This would mean that the way of thinking and analysing the HARM condition scenario is probably different from the way of thinking and analysing the HELP scenario. In the HARM scenario, one normative order, which Waleszczyński calls moral, dominates, while in the HELP condition scenario normative orders "compete" with one another.

As to the question asked at the onset of this article, namely, whether the way of thinking about the intentionality of causing a side effect in morally negative situations affects the way of thinking about the intentionality of causing a side effect in morally positive situations, or vice versa, the answer could be as follows. It is very likely that the way of thinking and analysing each of the scenarios depends on the normative order from the perspective of which each particular scenario or sequence of scenarios is considered. At the same time, the results suggest that it is moral normativity that decides the stability of the Knobe effect. Nevertheless, more in-depth empirical and theoretical studies are required in order to analyse the problems discussed in this article more thoroughly.

References

1. Adams, F., and A. Steadman. Intentional Action and Moral Considerations: Still Pragmatic, *Analysis* 64 (3), 2004, pp. 268-76.
2. Adams, F., and A. Steadman. Intentional Action in Ordinary Language: Core Concept or Pragmatic Understanding? *Analysis* 64 (2), 2004, pp. 173-81.
3. Butler, R. J. Report on Analysis "Problem" no. 16. *Analysis* 38 (3), 1978, pp. 113-4.
4. Guglielmo, S., and B. F. Malle. Can Unintended Side Effects Be Intentional? Resolving a Controversy Over Intentionality and Morality, *Personality and Social Psychology Bulletin* 36 (12), 2010, pp. 1635-47.
5. Harman, G. Practical Reasoning, *The Review of Metaphysics* 29 (3), 1976, pp. 431-63.
6. Hindriks, F. Control, Intentional Action, and Moral Responsibility, *Philosophical Psychology* 24 (6), 2011, pp. 787-801.
7. Hindriks, F. Intentional Action and the Praise Blame Asymmetry, *Philosophical Quarterly* 58 (233), 2008, pp. 630-41.

8. Holton, R. Norms and the Knobe Effect, *Analysis* 70 (3), 2010, pp. 417-24.
9. Knobe, J. Intention, Intentional Action and Moral Considerations, *Analysis* 64 (2), 2004, pp. 181-7.
10. Knobe, J. Intentional Action and Side Effects in Ordinary Language, *Analysis* 63 (3), 2003, pp. 190-4.
11. Knobe, J. The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology, *Philosophical Studies* 130 (2), 2006, pp. 203-31.
12. Leslie, A. M., J. Knobe, and A. Cohen. Acting Intentionally and the Side-Effect Effect, *Psychological Science* 17 (5), 2006, pp. 421-7.
13. Malle, B. F., and S. E. Nelson. Judging Mens Rea: The Tension Between Folk Concepts and Legal Concepts of Intentionality, *Behavioral Sciences and the Law* 21, 2003, pp. 563-80.
14. Mele, A., and S. Sverdlik. Intention, International Action, and Moral Responsibility, *Philosophical Studies* 82 (3), 1996, pp. 265-87.
15. Nadelhoffer, T. Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality, *Philosophical Explorations* 9 (2), 2006, pp. 203-19.
16. Nadelhoffer, T. Desire, Foresight, Intentions, and Intentional Actions: Probing Folk Intuitions, *Journal of Cognition and Culture* 6 (1-2), 2006, pp. 133-57.
17. Nadelhoffer, T. On Praise, Side Effects, and Folk Ascriptions of Intentionality, *Journal of Theoretical and Philosophical Psychology* 24 (2), 2004, pp. 196-213.
18. Nado, J. Effects of Moral Cognition on Judgments of Intentionality, *British Journal for the Philosophy of Science* 59 (4), 2008, pp. 709-31.
19. Nichols, S., and J. Ulatowski. Intuitions and Individual Differences: The Knobe Effect Revisited, *Mind and Language* 22 (4), 2007, pp. 346-65.
20. Paprzycka, K. Poznań Studies in the Philosophy of the Sciences and the Humanities. In A. Kuźniar, and J. Odrowąż-Sypniewska (eds.), *The Sciences*, Leiden, Boston: Brill Rodopi, 2016, pp. 204-33.
21. Paprzycka, K. The Omissions Account of the Knobe Effect and the Asymmetry Challenge, *Mind and Language* 30 (5), 2015, pp. 550-71.
22. Piekarski, M. One or Many Normativities? *Studia Philosophiae Christianae* 54 (1), 2018, pp. 5-24.
23. Piekarski, M. Two Arguments Supporting the Thesis About the Predictive Nature of Reasons for Action, *Studia Philosophiae Christianae* 54 (1), 2018, pp. 93-119.
24. Waleszczyński, A. Dwa porządki normatywne. Komentarz do dyskusji o intencjonalności działania, *AVANT. The Journal of the Philosophical-Interdisciplinary Vanguard* 8 (3), 2017, pp. 119-28.