University of Groningen

Control, intentional action, and moral responsibility

Hindriks, Frank

Published in:
Philosophical Psychology

DOI:
[10.1080/09515089.2011.562647](https://doi.org/10.1080/09515089.2011.562647)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Hindriks, F. (2011). Control, intentional action, and moral responsibility. *Philosophical Psychology*, 24(6), 787-801. <https://doi.org/10.1080/09515089.2011.562647>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Control, intentional action, and moral responsibility

Frank Hindriks

Skill or control is commonly regarded as a necessary condition for intentional action. This received wisdom is challenged by experiments conducted by Joshua Knobe and Thomas Nadelhoffer, which suggest that moral considerations sometimes trump considerations of skill and control. I argue that this effect (as well as the Knobe effect) can be explained in terms of the role normative reasons play in the concept of intentional action. This explanation has significant advantages over its rivals. It involves at most a conservative extension rather than a radical revision of what we tend to believe about intentional action, and it fits better with the way we conceive of the relation between intentional action and moral responsibility.

Keywords: Blameworthiness; Control; Intentional Action; Knobe Effect; Luck; Moral Responsibility; Normative Reason; Skill

1. Introduction

In a series of experiments, Joshua Knobe (2003a, 2003b, 2004) has discovered that our judgments about intentional action are sensitive to normative issues. The experiments concern the responses of people to agents who are indifferent regarding the effects of their actions, on the one hand, and the responses to agents who lack control over the effects on the other. Consider first the responses of people to cases involving agents who do not care about the harmful or beneficial consequences of their actions, to which I shall refer as “not-care cases.” It turns out that whether they regard bringing about an effect as an intentional action in such a situation depends on whether the effect is beneficial or harmful. More specifically, when people consider an indifferent agent, they tend to regard his bringing about an effect as an intentional action when it is harmful but not when it is beneficial. Furthermore, they are inclined to blame the agent in the former situation, but hardly praise him in the latter. Other experiments that Knobe (2003a, 2003b, 2004) has conducted pertain to

---

Frank Hindriks is Assistant Professor at the Faculty of Philosophy of the University of Groningen, the Netherlands.

Correspondence to: Frank Hindriks, Faculty of Philosophy, University of Groningen, Oude Boteringestraat 52, 9712 GL Groningen, The Netherlands. Email: f.a.hindriks@rug.nl

an agent who succeeds in doing something over which he has little or no control (see also Nadelhoffer, 2004a, 2005). Usually such doings are not intentional actions, but this may be different when the doing is normatively significant. In these experiments people deem the bringing about of an effect over which the agent has little or no control as intentional, both when they regard it as blameworthy and when they regard it as praiseworthy. Whereas the not-care cases have received a lot of attention in the literature (see Feltz, 2007; Nadelhoffer, 2010, for reviews), the lack-of-control cases have hardly been addressed. In this paper I propose an explanation of the lack-of-control cases that complements the explanation of the not-care cases I have provided elsewhere (Hindriks, 2008).¹

At the heart of my proposal lies what I call the “Normative Reason Hypothesis.” According to this hypothesis, normative considerations can make a difference as to whether or not a doing is an intentional action. The normative reason hypothesis can account for both the not-care cases and the lack-of-control cases. Ordinarily, bringing about an effect with respect to which one is indifferent is not an intentional action. But it is an intentional action when the effect is harmful and constitutes a normative consideration that counts against the intended action. Generally, lacking control over an effect of one’s action implies that the agent does not bring it about intentionally. I take the lack-of-control cases to reveal that such doings can be intentional actions when normative considerations count for or against trying to bring them about (in section 3, I specify the conditions under which this is the case). So, I take Knobe’s data to reveal something important about the notion of intentional action, and use them as input for a (partial) account of that notion. The account illuminates the role that normative considerations play in relation to intentional action. Because of the role normative reasons play in it, I call this the “Normative Reasons Account of Intentional Action,” or NORIA.²

Many believe that the experimental findings reveal biases in the formation of judgments about intentional action. They are particularly surprised by the fact that whether an effect of an action is beneficial or harmful makes a difference to attributions of intentional agency. The most popular strategy for explaining away this asymmetry is to appeal to the parallel asymmetry in our praise and blame judgments. The claim made is that judgments about praise- and/or blameworthiness illegitimately feed into our judgments about intentional action (Malle & Nelson, 2003; Mele, 2001; Nadelhoffer, 2004a, 2005; Nado, 2008).³ This claim runs counter to the commonsense conception of the relation between intentional action and moral responsibility. We ordinarily regard judgments about intentional action as input for judgments about moral responsibility. NORIA reveals that the experimental findings do not require us to abandon this idea. I regard the fact that it supports our commonsense conception of the relation between intentional action and moral responsibility as an important (non-experimental) reason for accepting NORIA (see sections 1, 2 and 4). So, I take NORIA to be supported by a combination of empirical and conceptual arguments.

Section 1 introduces the normative reason hypothesis and explains how it serves to account for the not-care cases. In section 2, I defend the main claim of

this paper: the Normative Reason Hypothesis can also be used to account for the lack-of-control cases. In section 3, I formulate NORIA, which also accommodates some more mundane cases of intentional action, and thereby offer a more complete (albeit still partial) account of intentional action. I go on to compare my explanation of the lack-of-control cases to its rivals in section 4.

2. Intentional Action, Moral Responsibility, and the Knobe Effect

Consider the chairman of a company who sets out to maximize profits. He is alerted to the fact that the business strategy that he intends to implement has an effect on the environment. In response, the chairman says that he does not care about the environment. In one version of the scenario the effect that the business strategy has on the environment is harmful, in another the effect is beneficial. Experimental findings show that most people say the chairman intentionally harms the environment, while most people say he does not intentionally benefit the environment (Knobe, 2003a). This asymmetry in our judgments about intentional action is known as the “Knobe effect.” The asymmetry also surfaces in other cases in which the agent is indifferent with respect to the effects of his action, i.e., in other not-care cases. Another scenario used in this study concerns a lieutenant who cares only about taking control of Thompson Hill. He thereby endangers the lives of his soldiers in one version of the scenario and shields their lives in another. Only in the former case is the effect that taking control of Thompson Hill has on the lives of his soldiers regarded as an intentional action.

I believe that the Knobe effect is best accounted for by embracing the idea that normative considerations can make a difference as to whether or not a doing is an intentional action. This is the Normative Reason Hypothesis I mentioned in the introduction. Before bringing this hypothesis to bear on the Knobe effect, we need to know what normative considerations or normative reasons are. Normative reasons are considerations that an agent should take into account when deliberating about an action. Considerations that actually influence an agent’s decision are motivating reasons. As normative reasons can but need not influence an agent’s decision, only some normative considerations are also motivating reasons (for particular agents at certain times). This is why normative reasons can be used to account for cases in which an agent is said to act intentionally, even though he is indifferent with respect to the effect he brings about. The thing to appreciate is that, whereas not caring about a normatively neutral effect implies that it is not brought about intentionally, this does not always hold for normatively significant effects.

How can the idea that normative reasons bear on intentional action explain the Knobe effect? In my (2008) I argue that a doing can be an intentional action due to the fact that the agent ignores a normative reason that counts against her intended action, and that this does not hold for normative considerations that count in favor of the action. In both kinds of cases, the agent does not care about a particular normatively significant effect of the action he envisages, such as the harm or benefit

to the environment. He thereby ignores a normative requirement. Given the moral importance of the effect, he should ascribe significance to it. When the effect is negative we flag this by judging that he brought it about intentionally. When the effect is positive we do not do so.⁴

This normative reason explanation of the Knobe effect differs from the one that Knobe has provided in important respects. Knobe (2006) has argued that it is the valence of the effect that explains the asymmetry in our judgments. When the effect is bad we conclude that the agent brings it about intentionally, whereas we do not do so when the effect is good. The thing to see is that the valence of the effect does not depend on the mindset of the acting agent. This implies that Knobe has to abandon the widely accepted idea that the notion of intentional action concerns the frame of mind with which an agent acts.⁵ The normative reason explanation preserves this idea by positing the agent's indifference as the key explanatory factor. An agent who is indifferent with respect to the normative significance of an effect of his intended action is aware of a normative reason that bears on that action but ignores it. The role that awareness plays in the proposed explanation secures the connection with the agent's mindset.

A related advantage of the normative reason explanation is that it can be used to make sense of the asymmetry in our judgments about intentional action. Again, indifference plays a central role here. As it turns out, our judgments about moral responsibility exhibit an asymmetry similar to that of the Knobe effect. We blame people in situations in which they ignore negative normative reasons, but we do not praise them when they fail to let their behavior be guided by positive normative reasons (Scanlon, 1998, p. 271; Wolf, 1990, p. 80). The underlying idea is that praise requires being motivated by what is good—an agent is praiseworthy only when she does what she does for the right reasons—whereas blame does not require being motivated by what is bad. An agent can also be blameworthy due to not being sufficiently motivated to avoid what is bad.⁶ When an agent is motivated by what is good or by what is bad, this will be reflected in her intention, because she acts at least in part in order to achieve that which is good or that which is bad. This is not the case when an agent is indifferent about an effect of her action.⁷ Even so, it serves a purpose in qualifying the agent's motivation as being such as to (defeasibly) warrant blame in such situations. This is what we do when we characterize the agent's bringing about of the effect as an intentional action (even though it is not intended). It would not make sense to do the same in positive cases, because we do not praise agents whose motivation is deficient. The upshot is that, if the notion of intentional action is to be useful for arriving at judgments concerning moral responsibility, it needs to involve an asymmetry that runs parallel to the asymmetry in our judgments about praise and blame. It makes a lot of sense, from the perspective of our concept of moral responsibility, to have a concept for marking a flaw in someone's motivation only when a negative rather than a positive effect is at issue. If the normative reason explanation of the not-care cases is correct, this is our concept of intentional action.

As mentioned in the introduction, Knobe (2003a, 2003b, 2004) and Nadelhoffer (2004a, 2005) have also investigated cases in which the agent does not fully control what he does, which I call “lack-of-control cases.” These differ from the scenarios involved in the Knobe effect in that the agent does care about the consequences of his actions. To give an example, one of the experiments Knobe conducted concerns someone who desperately wants to win a rifle contest, even though he is not very good at using his rifle. This does not mean, however, that the preceding is irrelevant to lack-of-control cases. The core insight underlying the explanation of the Knobe effect is that normative considerations can make a difference as to whether someone acts intentionally. It may well be that this also holds in some cases in which an agent does care about a normatively significant effect of his action. The next section is devoted to the question how the Normative Reason Hypothesis can be used to account for judgments of intentional action in the absence of (full) control. Having presented more of the details of my account of intentional action in section 3, I argue in section 4 that the advantages of the normative reason hypothesis discussed here extend to the proposed explanation of the lack-of-control cases.

3. Lack-of-Control Cases

Suppose you throw a fair die hoping to roll a six. Whether you do so is a matter of sheer luck. It is widely accepted that doings such as this, in which the agent is engaged only (or mostly) by luck, are not intentional actions. The experiments Knobe and Nadelhoffer conducted concerning such lack-of-control cases provide reason to reconsider the relation between luck and intentional action. The most striking finding is that our judgments concerning intentional action in lack-of-control cases are sensitive to normative considerations. Nadelhoffer concludes that “considerations of luck and skill can sometimes be trumped by moral considerations” (2004a, p. 283). Let me review some of the experiments on which this conclusion is based.

Knobe (2003b) considers an agent, Jake, who uses his rifle to shoot at a target. In some scenarios he is an expert marksman, while in others he is not very good at using his rifle. In the non-moral scenarios, Jake desperately wants to win a rifle contest. In the moral scenarios, Jake stands to inherit a lot of money when his aunt dies, and tries to kill her using his rifle. In morally neutral cases, the percentage of subjects who regard the bringing about of the effect as an intentional action when Jake lacks (the skill and hence) control goes down from 79 to 28. In the morally charged cases involving Jake’s aunt, the percentages are 95 and 76, respectively. What is important for my current purposes is the fact that there is a marked contrast between the non-moral and the moral lack-of-control case. Whereas only 28% of the participants in Knobe’s experiment conclude that Jake intentionally hits the bull’s eye in the rifle contest, 76% judge that he kills his aunt intentionally. Whereas intentionality ratings go down dramatically in the morally neutral version when the agent lacks control, a large majority still regards the doing as an intentional action when the effect is normatively significant, even when the agent lacks control.

Nadelhoffer (2005) argues that Knobe's study is not as methodical as it should have been, and sets out to investigate lack-of-control cases in a more systematic way. Knobe already observes that "the 'no-skill' conditions lie somewhere in the gray area between total skill and total luck" (2003b, p. 313). Nadelhoffer suggests considering cases where the agent has no control over the outcome whatsoever, as is the case when someone throws a fair die. He also notes that the questions Knobe asks in relation to the relevant scenarios are asymmetric: in the morally neutral cases subjects are asked whether Jake hit the bull's eye intentionally, while in the morally charged cases the question asked is not whether Jake shoots his aunt intentionally, but whether he kills her intentionally. Killing the aunt and winning the rifle contest are ends for which attempted shootings are means. It would have been better had Knobe formulated his questions consistently in terms of either means or ends. The scenarios Nadelhoffer considers do not suffer from these flaws, and carry a promise of providing better evidence for the claims made. Nadelhoffer has indeed replicated Knobe's findings.

Two of the cases that Nadelhoffer (2005) considers concern a nuclear reactor. In one version the agent, Fred, tries to prevent an explosion. The only way he can do so is by punching a certain ten-digit code into a computer. However, Fred has no idea what the code is. By sheer luck the code he types in turns out to be correct. Thus he prevents a nuclear meltdown and saves thousands of people. In the other version, Fred is intent on causing a nuclear meltdown. Again he ends up entering the correct code by sheer luck. Rather than saving them, he ends up killing thousands of people. It turns out that a majority of the subjects deem Fred's action of preventing or causing the explosion to be intentional (73% and 83%, respectively). The participants also assign to Fred a fair amount of praise and blame for his actions (the average ratings were 4.0 and 5.31 on a 6-point scale, respectively).

The lack-of-control cases are classified as intentional actions exactly when the agent's action is normatively significant. In order to determine how this observation can be used to explain the findings, it will again be useful to consider moral responsibility. Suppose Fred is indeed praiseworthy when he prevents a nuclear meltdown and blameworthy when he causes one. This will be due to the way he responds to the normative reasons that bear on his action. When the effect is positive, the fact that the effect might obtain provides a normative reason in favor of trying to bring it about. When the effect is negative, this should count against trying to perform the action. Supposing these reasons are decisive, the way Fred responds to them supports the praise and blame judgments, as he (fails to) respond(s) appropriately to the relevant normative reason in the particular case.

What would the notion of intentional action have to be like if it were to be conducive to forming these judgments about moral responsibility? In section 1, we saw that being motivated by good (bad) provides support for praising (blaming) the agent, and that blame can also be due to a deficiency in someone's motivation to avoid what is bad. When he attempts to prevent the nuclear meltdown, Fred has the morally appropriate motivation. It would be useful to have a notion that flags this. The idea underlying the Normative Reason Hypothesis is that this is the notion of

intentional action. When Fred tries to cause a nuclear meltdown, his motivation is morally deficient and warrants blame. Again, the notion of intentional action should be such as to imply that the meltdown is brought about intentionally in order to serve its purpose in relation to moral responsibility. This is the case if the notion of intentional action is sensitive to the way people respond to normative reasons in the same way as the notion of moral responsibility. In light of this I propose that, even though a lack of control over a behavior normally implies that it is not an intentional action, this need not be the case when the behavior is normatively significant (the exact conditions under which such normatively significant behaviors are intentional actions are specified in section 4). This idea provides the core of the normative reason explanation of the lack-of-control cases.

The proposed explanation reveals something important about the relation between intentional action and moral responsibility. Often, characterizing a morally significant action as intentional amounts to qualifying the agent's motivation as (defeasibly) warranting praise or blame. In morally neutral cases, not having the requisite skill to bring an effect about implies that the agent did not bring it about intentionally. This is different when the effect is normatively significant, because in such cases the way in which an agent (fails to) respond(s) to the normative considerations that bear on his action may warrant praise or blame. In section 1, we saw that something similar holds for cases in which an agent is unconcerned with an effect of his action. The upshot is that neither the not-care cases nor the lack-of-control cases necessitate a revision of the commonsense conception of the relation between the two notions at issue (this in spite of what those who take the findings to reveal biases in the formation of judgments about intentional action believe; see section 4 for further discussion). What is more, a careful investigation of this relation provides the key to an explanation of these experimental findings.

4. The Normative Reasons Account of Intentional Action

Thus far I have been concerned with special cases of intentional action, cases where the agent does not care about a particular effect of her action as well as cases where she lacks the control to bring it about. In more mundane cases of intentional action the agent does care about what she does and has control over it. In order to put the accounts presented in the previous sections in perspective, it is useful to provide a more complete picture of intentional action, and also to discuss some other mindsets that can make a doing an intentional action. If the not-care cases and the lack-of-control cases are indeed genuine instances of intentional action, an agent can, for instance, perform an action intentionally without intending to perform that action. As I shall discuss shortly, this also holds for some of the more mundane cases. In addition to providing more context, I add more detail. I develop the accounts of the not-care cases and the lack-of-control cases into proposals for sufficient conditions for intentional action (that may require further refinement in light of new empirical evidence). Together with conditions that account for the more mundane

cases, this results in a (still partial) account of intentional action that does justice to the roles that both motivating and normative reasons play in that concept. This is the “Normative Reasons account of Intentional Action,” or NORIA.

In the cases I propose to consider now, you are in control and you care about the effect of your intended action, either because you want to or because you do not want to bring it about. Suppose you regularly go for a swim. Your primary motivation is just that you like swimming. However, you are also partly motivated to go swimming because of the positive consequences for your health—let’s say because you expect to lose weight. This consideration is strong enough to (positively) affect the frequency with which you swim. In such a situation you lose weight intentionally. Next, imagine that things are very busy at work. You decide to work late again, even though you know your spouse will be disappointed, because this means you will spend less time with him or her. You would rather not disappoint your spouse, but the issues you have to attend to at work are so important that you decide to give them priority. Under these conditions you disappoint your spouse intentionally (even though you do not intend to do so).⁸

What these two examples share is that you intend to perform a particular action, and you expect to bring about a certain effect by doing so. In the first example the expected effect functions as a reason for performing the intended action. You intend to swim in part because you expect to lose weight by doing so. In the second example you have formed your intention to work late despite the effect you expect this to have on your spouse, and you treat that effect as a consideration that counts against performing the intended action. Whereas you do want to lose weight, you do not want to disappoint your spouse. In both cases you care about the effect. In the former you ascribe positive significance to it, while you ascribe negative significance to the effect in the latter. These two kinds of cases of intentional action can be captured as follows:

An agent *S* φ s intentionally if she intends to ψ , φ s by ψ ing, and:

- (a) *S* expects to φ by ψ ing, and intends to ψ because she expects thereby to φ , or;
- (b) *S* expects to φ by ψ ing, and ψ s in spite of the fact that she does not want to φ .

Here φ is an action that involves an effect generated by your intended action.⁹ So, an example of φ ing by ψ ing is disappointing your wife by working late. An assumption that figures in the background of this account is that it is rational for *S* to expect to φ by ψ ing, and that this is in part because she has the requisite control over her φ ing by ψ ing.

Condition (a) is not particularly controversial (if at all). Knobe and Mendlow (2004) conducted a study that provides confirmation for condition (b). The scenario they tested involves a president of a computer corporation, Susan, who implements a business program that is expected to lead to an increase in sales in Massachusetts and a decrease in sales in New Jersey. She comments on these effects of the program as follows: “according to my calculations, the losses we sustain in New Jersey should be a little bit smaller than the gains we make in Massachusetts. I guess the best course of action would be to approve the program” (Knobe & Mendlow, 2004, p. 257). It is

rather plausible to interpret this passage as implying that Susan likes the increase in sales, dislikes the decrease, and favors the program because it leads to an overall increase. This would mean she does not want sales in New Jersey to decrease, and that she acts in spite of this effect. So all the elements of condition (b) are satisfied. As it turns out, 75% of the participants in the study take Susan to decrease the sales in New Jersey intentionally.¹⁰

The not-care cases involved in the Knobe effect can be accounted for by introducing a condition according to which the agent is not at all concerned with the fact that her intended action has a certain effect. At the same time, the effect should be normatively significant in the sense that the agent ought to ascribe negative significance to it. The following condition captures intentional actions that have these features:

An agent *S* φ s intentionally if she intends to ψ , φ s by ψ ing, and:
(c) *S* expects to φ by ψ ing, does not care about her φ ing by ψ ing, and ψ s in spite of the fact that she believes her expected φ -ing constitutes a normative reason against her ψ ing.

This condition can explain why the chairman's harming the environment discussed in section 1 is considered an intentional action.¹¹ What has to be assumed for the explanation to go through is that the chairman is aware that, even though he does not care about the environment, he ought to ascribe negative significance to harming the environment. This does not appear to be a farfetched assumption. Recall that the Knobe effect is an asymmetry in our judgments about intentional action: the chairman's helping the environment is not regarded as an intentional action. This is accounted for by requiring that there not be an analogous condition that involves positive normative reasons, or considerations an agent should take to reflect positively on the intended action.

I argued in section 2 that the Normative Reason Hypothesis, which I have used to account for the not-care cases, can also be used to account for the lack-of-control cases, which provide the focus of this paper. In light of this it is natural that the notion of a "normative reason" also features in the condition that accounts for the lack-of-control cases. Such cases cannot involve a rational expectation to bring about the relevant effect. However, a rational agent can hope to do so. Adding a condition to the account that differs in this respect from the preceding one serves to account for the lack-of-control cases presented by Knobe and Nadelhoffer:

An agent *S* φ s intentionally if she intends to ψ , φ s by ψ ing, and:
(d) *S* hopes to φ by ψ ing, and ψ s because (in spite of the fact that) she believes the φ -ing constitutes a normative reason in favor of (against) her ψ ing.

The fact that an agent who meets the requirements of condition (d) hopes to φ implies that the agent is not indifferent with respect to φ ing, but wants to φ . Note that condition (d) is symmetrical in that it allows for both positive and negative normative reasons.

It is quite plausible that the requirements of condition (d) are satisfied in Nadelhoffer's nuclear reactor examples, which were discussed in the previous section.

The vignettes that were used include the sentence: “vividly aware that the odds against typing in the correct code are astronomical, Fred decides to give it a try” (Nadelhoffer, 2005, pp. 347–348). Presumably Fred gives it a try under this circumstance because he hopes he might succeed. (Hope figures explicitly in the scenario that Ronald Butler, 1977, proposed, which inspired many of Nadelhoffer’s examples.) The same holds for Knobe’s example concerning Jake, who wants to kill his aunt even though he is not very good at using his rifle.

Many philosophers of action accept that conditions (a) and (b), which concern motivating reasons, are sufficient conditions for intentional action (see note 8 for references). The conditions that are distinctive of NORIA are conditions (c) and (d), the conditions that involve normative reasons (in condition d the normative reason also features as a motivating reason). These conditions have been designed to account for the not-care cases and the lack-of-control cases respectively. They are applications of the Normative Reason Hypothesis discussed in sections 1 and 2. Thus NORIA makes precise how normative reasons bear on intentional action.

5. Putting the Normative Reason Hypothesis in Perspective

If the Normative Reason Hypothesis in general and condition (d) of NORIA in particular, are correct, the lack-of-control cases can be explained in terms of the normative reasons that bear on those cases. In this section I compare this explanation to its rivals. Knobe (2003b) considers what he calls the “Blame Hypothesis” (see also Nadelhoffer, 2004b, 2004c). According to the Blame Hypothesis, behavior is regarded as intentional if it is “especially blameworthy,” even when the agent had little or no control over it (Knobe, 2003b, p. 320). As Knobe recognizes, the blame hypothesis fails to account for lack-of-control cases involving behavior that is morally good. The lack-of-control cases are symmetrical in that the behavior is regarded as intentional in both good and bad cases. The Blame Hypothesis cannot account for this symmetry.¹²

The natural move to make in response to this problem is to consider what I call the “Praise/Blame Hypothesis.” According to this hypothesis, a lack of control only implies the absence of intentional agency in morally neutral cases. In cases in which the agent is blame- or praiseworthy, a lack of control is consistent with acting intentionally. Nadelhoffer (2005) appears to support this hypothesis. He argues that “moral considerations sometimes trump considerations of skill, luck, and control” (Nadelhoffer, 2005, p. 350), and the moral considerations that figure most prominently in his paper are praise and blame. He also maintains that the results of his experiments “show that when an agent’s actions are *sufficiently praiseworthy or blameworthy*, people are less interested in whether or not the action was the result of skill or luck: (Nadelhoffer, 2005, p. 349). The Praise/Blame Hypothesis fits the data. However, it is not the only existing hypothesis that does so.

Knobe (2003b) has proposed what he calls the “Dissociation Hypothesis.” As Knobe uses the term, dissociation concerns the relation between the agent’s praise- or

blameworthiness and her (lack of) skills: “people dissociate the agent from the behavior when they conclude that the agent deserves less praise or blame because of her lack of skill” (2003b, p. 318). The idea underlying this notion of “dissociation” is that a lack of skill often, though not always, diminishes the extent to which people praise or blame the agent. If it does not, the relevant behavior constitutes an intentional action. If, however, the lack of skill does diminish the extent of praise or blame, the behavior is not an intentional action. The point of reference is the extent to which the agent is praised or blamed for the behavior in a situation where she does have full control over it. The agent engages in the behavior intentionally if there is little or no difference in the extent to which she is praised or blamed for it when she lacks control over it as compared to when she has control over it. (As the average praise and blame ratings are high, this appears to hold for Nadelhoffer’s nuclear reactor examples.) In sum, the Dissociation Hypothesis says that a behavior over which the agent had little or no control is nevertheless an intentional action, if the lack of (full) control does not significantly affect the extent to which the agent is praised or blamed. The less people dissociate the agent from the behavior, the more willing they will be to consider that action intentional.¹³

How does my normative reason explanation of the lack-of-control cases compare to these rival explanations? When discussing Knobe’s valence explanation in section 1, I argued that the normative reason explanation allows us to preserve the widely accepted idea that intentional action concerns the frame of mind with which an agent acts, and that it does not conflict with the commonsense conception of the relation between intentional action and moral responsibility. Similar considerations apply here. Let me elaborate. Both the Praise/Blame Hypothesis and the Dissociation Hypothesis appeal to the judgment of people other than the agent. This conflicts with the idea that whether an agent acts intentionally is a matter of the frame of mind with which she acts (see Malle & Knobe, 1997, for empirical support). As it appeals to beliefs the agent has about the normative reasons that bear on the action she intends to perform, the normative reason explanation allows us to preserve this idea. Thus it fits better with our intuitions about intentional action. Rather than a radical revision, my proposal is a conservative extension of commonly accepted ideas about intentional action.

Another advantage of the normative reason hypothesis is that it sheds light on our praise and blame judgments. Earlier we saw that acting on (against) the normative reason’s bearing on an action provides support for praise (blame). This is what the agents who satisfy the requirements of condition (d) of NORIA do. In sections 1 and 2, I argued that characterizing a morally significant action as intentional often amounts to qualifying the agent’s motivation as (defeasibly) warranting praise or blame. Thus (only) the Normative Reason Hypothesis allows us to retain the idea that judgments about intentional action can provide useful input for judgments about moral responsibility or praise and blame.¹⁴

These two advantages are rather similar to the advantages that the normative reason explanation of the not-care cases enjoys over Knobe’s valence explanation. The third advantage that the normative reason explanation of the not-care cases

enjoys relative to the Praise/Blame Hypothesis and the Dissociation Hypothesis is independent of these. The normative reason explanation serves to make sense of why a lack of control does not always rule out intentional agency. In many cases a lack of control blocks the ascription of blame. Consider an employee of a factory who operates a machine that breaks down and injures a colleague of his in the process. He had taken all the relevant safety precautions, so the event is truly an accident. We do not blame him, because as Sher explains, “when someone performs a wrong act over which he lacks control, the wrongness of what he does cannot be attributed to his failure to respond to the relevant moral reason” (2006, p. 58). Now, it is important to see that this does not hold for the lack-of-control cases discussed above. In these cases the fact that the relevant effect might come about provides a powerful normative reason (not) to perform the intended action in the first place. Knobe’s example nicely illustrates this point. The chance that Jake will end up killing his aunt may be small given that he is bad at using his rifle. Even so, he should not even try to shoot her exactly because he might end up killing her in this way.

This explanation reveals that control is, after all, in a sense a necessary condition for intentional action. What happens in cases where the effect that might result from a particular action is normatively significant, is that the attention shifts from the fact that one does not have control over the effect the action might generate, to the fact that one engages in that action in the first place, as one thereby incurs the risk of generating that effect. An agent does have control over whether she engages in that action.

Thus, in addition to preserving core intuitions about intentional action and its relation to moral responsibility, the Normative Reason Hypothesis has more explanatory power than the Praise/Blame Hypothesis and the Dissociation Hypothesis. These hypotheses take for granted that in some cases a lack of control does not decrease the extent to which we praise or blame someone. This, however, is something that itself cries out for explanation. Zooming in on the role normative reasons play in relation to cases in which control is absent serves to explain this, and provides the basis for a systematic, well-motivated account of our judgments about intentional agency. All in all, I conclude that there are good grounds for preferring the Normative Reason Hypothesis to its rivals.

Acknowledgements

I gratefully acknowledge helpful comments from Joshua Knobe.

Notes

- [1] See Lanteri (2009) for a critique.
- [2] NORIA is a competence account rather than a performance (error) account (see Nado, 2008, for a useful discussion of this distinction).
- [3] Since then, Mele (2003) has rejected this idea in light of new experimental findings, presented in Knobe (2003a).

- [4] Wible (2009) also suggests that the fact that the agent does not care, even though he should, affects our judgments about intentionality. Contrasting his account to Knobe's, he maintains that "it is the attitude of the chairman to the side effect that matters more than the goodness or badness of the side effect itself" (2009, p. 176).
- [5] Malle and Knobe (1997) provide empirical support for this idea when they identify belief, desire, intention, awareness, and skill as key components of our concept of intentional action.
- [6] Arpaly's distinction between good will, ill will, and deficiency in good will is illuminating: "if good will is responsiveness to moral reasons, deficiency in good will is insufficient responsiveness to moral reasons, and ill will is responsiveness to sinister reasons—reasons that, in their essence conflict with morality" (2002, p. 231). On her view, praiseworthiness requires a good will, while blameworthiness does not require an ill will, but is consistent with a deficiency in good will.
- [7] Stocker captures these insights nicely when he claims that "an act is not praiseworthy nor, therefore, is it supererogatory unless done with a good intention. . . . A blameworthy act need not be done with a bad intention" (1973, p. 60).
- [8] See Bratman (1987), Enç (2003), and Harman (1976), for similar examples. See below for a discussion of empirical evidence.
- [9] In the formulation of conditions of intentional action here and elsewhere in the paper I abstract from the problem of deviant causal chains. As is well known, an agent ϕ s intentionally by ψ ing only if his ψ ing causes his ϕ ing "in the right way." See Enç (2003), and Mele and Moser (1994) for insightful explications of this qualification.
- [10] Some studies suggest things are different when the effect is normatively significant. Phelan and Sarkissian (2008) discuss a case in which a city planner implements a project in order to solve the city's pollution problem. The program also increases joblessness, which the city planner regrets. It turns out that only 29% take him to increase joblessness intentionally. Similarly, Mele and Cushman (2007) have run a study involving Ann, who fills in the pond next to her house because it is a breeding ground for mosquitoes. She takes herself to be responsible for this because she owns the lot. Some children look for frogs there, and they become sad because the pond is filled in. Only 28% say Ann makes the children sad intentionally. Mele and Cushman suggest this is due to the fact that "Ann may strike people as justified in filling in the pond at the expense of the kids' sadness" (2007, p. 194). So, perhaps an expected and undesired effect is not brought about intentionally when the agent has sufficient normative reason to bring it about. Further research is needed in order to determine exactly how condition (b) needs to be qualified. For now it is perhaps best to take it to apply to normatively neutral cases only.
- [11] This formulation differs from the one provided in Hindriks (2008) in that it explicitly stipulates the agent does not care about ϕ ing by ψ ing. This is because the evidence discussed in note 10 suggests that the agent's wanting or not wanting to ϕ when ψ is an action that has normative significance can make a difference regarding whether she ϕ s intentionally.
- [12] Knobe and Mendlow (2004) argue that the decreasing sales in New Jersey example discussed in section 3 provides evidence against the Blame Hypothesis, because Susan is said to act intentionally but not to be blameworthy (only 5% of the participants in their experiment attribute blame to her; Knobe & Mendlow, 2004, p. 257). As decreasing sales is in some non-moral sense bad, they take it to support the claim that the valence of the effect explains the data. Once a distinction is made between motivating and normative reasons, it turns out that the decreasing sales example does not provide evidence against the Blame Hypothesis, exactly because it is a normatively neutral case. It is accounted for by NORIA's condition (b), and not by (c). See Nado (2008, pp. 724–725) for a similar critique of Knobe and Mendlow.
- [13] The formulations Knobe uses include the following: "the more the agent is dissociated from his action, the more reluctant people will be to consider that action intentional," and "the

hypothesis says that people do not regard the behavior as intentional when there is a high degree of dissociation" (2003b, p. 319).

- [14] In my (2008) I argued that normative reasons are a common denominator of both intentional action and moral responsibility. Normative considerations are in some cases prior to intentional action, in that they figure as inputs into adequate judgments about intentional action. In contrast to what others have suggested, this does not preclude judgments about intentional action from figuring as inputs into reliable judgments about moral responsibility. The fact that these notions have a common denominator does not make one of them redundant either. The notion of intentional action is still useful because it serves to indicate whether an agent was motivated in such a way as to (defeasibly) warrant praise or blame for that action.

References

- Arpaly, N. (2002). Moral worth. *Journal of Philosophy*, 99, 223–245.
- Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.
- Butler, R. J. (1977). Analysis Competition "Problem" No. 16. *Analysis*, 37, 97.
- Enç, B. (2003). *How we act: Causes, reasons, and intentions*. Oxford: Oxford University Press.
- Feltz, A. (2007). The Knobe effect: A brief overview. *Journal of Mind and Behavior*, 28, 265–277.
- Harman, G. (1976). Practical reasoning. *Review of Metaphysics*, 29, 431–463.
- Hindriks, F. (2008). Intentional action and the praise-blame asymmetry. *Philosophical Quarterly*, 58, 630–641.
- Knobe, J. (2003a). Intentional action and side effects in ordinary language. *Analysis*, 63, 190–194.
- Knobe, J. (2003b). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, 16, 309–324.
- Knobe, J. (2004). Intention, intentional action and moral considerations. *Analysis*, 64, 181–187.
- Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies*, 130, 203–231.
- Knobe, J., & Mendlow, G. (2004). The good, the bad, and the blameworthy: Understanding the role of evaluative reasoning in folk psychology. *Journal of Theoretical and Philosophical Psychology*, 24, 252–258.
- Lanteri, A. (2009). Judgments of intentionality and moral worth: Experimental challenges to Hindriks. *Philosophical Quarterly*, 59, 713–720.
- Malle, B., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, 101–121.
- Malle, B., & Nelson, S. (2003). Judging mens rea: The tension between folk concepts and legal concepts of intentionality. *Behavioral Sciences and the Law*, 21, 563–580.
- Mele, A. (2001). Acting intentionally: Probing folk notions. In B. Malle, L. Moses, & D. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 27–43). Cambridge, MA: MIT Press.
- Mele, A. (2003). Intentional action: Controversies, data, and core hypotheses. *Philosophical Psychology*, 16, 325–340.
- Mele, A., & Cushman, F. (2007). Intentional action, folk judgments, and stories: Sorting things out. *Midwest Studies in Philosophy*, 31, 184–201.
- Mele, A., & Moser, P. (1994). Intentional actions. *Nous*, 28, 39–68.
- Nadelhoffer, T. (2004a). The Butler problem revisited. *Analysis*, 64, 277–284.
- Nadelhoffer, T. (2004b). Praise, side effects, and intentional action. *Journal of Theoretical and Philosophical Psychology*, 24, 196–213.
- Nadelhoffer, T. (2004c). Blame, badness, and intentional action: A reply to Knobe and Mendlow. *Journal of Theoretical and Philosophical Psychology*, 24, 259–269.

- Nadelhoffer, T. (2005). Skill, luck, control, and intentional action. *Philosophical Psychology*, 18, 341–352.
- Nadelhoffer, T. (2010). Experimental philosophy of action. In J. Aguilar, A. Buckareff, & K. Frankish (Eds.), *New waves in philosophy of action* (pp. 50–78). New York: Palgrave MacMillan.
- Nado, J. (2008). Effects of moral cognition on judgments of intentionality. *British Journal for the Philosophy of Science*, 59, 709–731.
- Phelan, M. T. & Sarkissian, H. (2008). The folk strike back; or, why you didn't do it intentionally, though it was bad and you knew it. *Philosophical Studies*, 138, 291–298.
- Scanlon, T. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Sher, G. (2006). *In praise of blame*. Oxford: Oxford University Press.
- Stocker, M. (1973). Act and agent evaluations. *Review of Metaphysics*, 27, 42–61.
- Wible, A. (2009). Knobe, side effects, and the morally good business. *Journal of Business Ethics*, 85, 173–178.
- Wolf, S. (1990). *Freedom within reason*. Oxford: Oxford University Press.