



# Experimental Philosophy, Ethnomethodology, and Intentional Action: A Textual Analysis of the Knobe Effect

Gustav Lymer<sup>1</sup> · Olle Blomberg<sup>2</sup>

Published online: 18 July 2019  
© The Author(s) 2019

## Abstract

In “Intentional action and side-effects in ordinary language” (2003), Joshua Knobe reported an asymmetry in test subjects’ responses to a question about intentionality: subjects are more likely to judge that a side effect of an agent’s intended action is intentional if they think the side effect is morally bad than if they think it is morally good. This result has been taken to suggest that the concept of intentionality is an inherently moral concept. In this paper, we draw attention to the fact that Knobe’s original interpretation of the results is based on an abstract rendering of the central scenario (the Chairman scenario) that is significantly different from the vignettes presented to the survey participants. In particular, the experimental vignettes involve temporal and social dimensions; they portray sequences of social actions involving an agent and an interlocutor, rather than a lone agent making a momentary decision in light of certain attitudes. Through textual analyses of a set of vignettes used to study the Knobe effect, drawing on ethnomethodology, conversation analysis, and discursive psychology, we show that there are many differences between the experimental conditions besides the moral valence of the side effect. In light of our textual analyses, we discuss vignette methodology in experimental philosophy and suggest an alternative interpretation of Knobe’s original experimental results. We also argue that experimental philosophy could benefit from considering research on naturally occurring social interaction as an alternative source of empirical findings for discussions of folk-psychological concepts.

**Keywords** Knobe effect · Intentional action · Textual analysis · Ethnomethodology · Experimental philosophy

---

✉ Gustav Lymer  
gustav.lymer@edu.su.se

<sup>1</sup> Department of Education, Stockholm University, Stockholm, Sweden

<sup>2</sup> Department of Philosophy, Lund University, Lund, Sweden

## Introduction

Is what an agent does intentionally simply what she intends to do and brings about as a result? According to the so-called Simple View, the answer is affirmative (Adams 1986). An alternative view holds that “there are cases where S intentionally does A without intending to do A, as long as doing A is foreseen and S is willing to accept A as a consequence of S’s action” (Adams and Steadman 2004: 173). Joshua Knobe (2003) reported survey results showing that side effects of intended actions are sometimes seen as intentional. Knobe also found that there is an asymmetry to such judgments: “people seem considerably more willing to say that a side-effect [of an agent’s intended action] was brought about intentionally when they regard that side-effect as bad than when they regard it as good”.<sup>1</sup> These survey results have since been replicated many times and the asymmetry has become known as “the Knobe effect” (for references, see, e.g., Knobe 2010; Cova 2016). The results suggest, first, that people at least implicitly reject the Simple View, since there are cases where unintended but foreseen side effects of an intended action are seen as intentional. Second, they suggest, according to Knobe (2003), that people’s everyday concept of intentional action is an inherently moral concept.

Adams and Steadman (2004) argue that it is doubtful that Knobe’s results actually tell us anything about “the core concept of intentional action”. Instead, they suggest that the results tell us something about people’s sensitivity to pragmatic aspects of “intentional language” and “intentional talk”. In this paper, we will argue for a similar view based on detailed textual analyses of some vignettes used to study the Knobe effect. Two initial observations frame our analyses: first, there is a textual discrepancy between the experimental vignettes and the theoretical discussion of the results; and secondly, the stories depicted in the vignettes have a *sequential* and often *conversational* character. Even though a number of authors have explored how features of the vignettes, including their wording, may contribute to intentionality judgments (Adams and Steadman 2004; Mele and Cushman 2007; Scaife and Webber 2013), the textual phenomena we highlight have not been acknowledged in prior discussions. We will argue, however, that they are significant.

In the next section, we present and contrast two renderings of one of the scenarios used in Knobe’s (2003) survey studies: a schematic rendering in the paper’s introduction and the vignette presented to the subjects doing the survey. In the “[Analytical Approach: Ethnomethodology and Conversation Analysis](#)” section, we then present the ethnomethodological and conversation analytic perspective that has informed our analyses. Then, in the “[Knobe’s Vignettes as Accounts of Action Sequences](#)” section, we provide a textual analysis of the vignettes used by Knobe (2003). We also briefly consider a variation of the study reported by Mele and Cushman (2007). In the “[Lone Agent Scenarios](#)” section, we go on to analyse two vignettes that do not

<sup>1</sup> This is Knobe’s original interpretation. He has formulated a more sophisticated interpretation to account for other experimental results in (Knobe 2010). However, our primary focus is the method of using vignettes in experimental philosophy rather than how to best explain existing experimental results, so to keep things relatively brief we will only discuss the original interpretation.

describe a conversation but rather a decision-making process carried out by a single agent (Nadelhoffer 2006; Knobe and Kelly 2009). In the “Discussion” section, we discuss what implications our textual analyses might have for how the Knobe effect should be understood. While we do not offer a full explanation of the effect, we show how the differences highlighted by our textual analyses are likely to influence subjects’ intentionality judgements. We also argue that, to explain the Knobe effect, we need not appeal to features external to the vignettes, such as the test subject’s moral evaluation of the side-effect.

## Decision Point and Sequence: Two Renderings of the Stimulus

Almost all empirical work on the Knobe effect and intentional action is based on survey studies.<sup>2</sup> Subjects read vignettes describing a sequence of events that ends with a choice made by a protagonist. This choice results in the successful execution of the intended action but also a foreseen good or bad side-effect. Knobe assumed that he could manipulate the moral valence of this side-effect by changing one or a few words in the vignettes, while keeping everything else in the story unchanged.

Knobe’s paper starts with the following schematic scenario:

The chairman of the board of a company has decided to implement a new program. He believes

(1) that the program will make a lot of money for his company

and

(2) that the program will also produce some other effect  $x$ .

But the chairman doesn’t care at all about effect  $x$ . His sole reason for implementing the new program is that he believes it will make a lot of money for the company. In the end, everything proceeds as anticipated: the program makes a lot of money for the company and also produces effect  $x$ . [...] The question I want to address here is: Shall we say that the chairman brought about this side-effect intentionally? (Knobe 2003: 190)

Knobe then reports that the subjects tended to judge that the chairman brought about  $x$  intentionally if  $x$  was the effect of harming the environment, but not if  $x$  was the effect of helping the environment. To Knobe, this suggested that the difference in moral valence of the side effect, from the subjects’ evaluative perspective, accounted for the asymmetry in responses. This is Knobe’s so-called “moral valence hypothesis” (MVH, [see Hindriks 2014]).

We call this schematic presentation in Knobe’s opening paragraphs the *decision-point rendering* of the Chairman scenario. As readers, we are presented with the states of knowledge and attitudes of a lone agent who has made a particular decision, which has certain described consequences. We are told that the decision was motivated by a desire for the main intended effect (profit) and means-end knowledge about what will

<sup>2</sup> Utikal and Fischbacher (2014) and Mizumoto (2018) are two exceptions.

bring this effect about. In addition, we are informed that, in making the decision, the agent knew of, but did not care about, an unspecified side effect X.

However, this is not the rendering presented to the subjects. Subjects were randomly assigned to read one of two vignettes, the *Harm* or the *Help* vignette. In the following, the differences between these two vignettes are marked with square brackets:

The vice-president of a company went to the chairman of the board and said, ‘We are thinking of starting a new program. It will help us increase profits, [but] [and] it will also [harm] [help] the environment.’

The chairman of the board answered, ‘I don’t care at all about [harming] [helping] the environment. I just want to make as much profit as I can. Let’s start the new program.’

They started the new program. Sure enough, the environment was [harmed] [helped]. (Knobe 2003: 191)

In this *sequential rendering*, the scenario unfolds as a sequence of social actions that leads to a decision by the agent. The vignettes involve temporal and interactional dimensions that are absent in the decision-point rendering. Each describes a decision-making process that unfolds as a dialogue and terminates in the decision, while, in the decision-point rendering, the decision is instead portrayed as already made. The interpretation of the survey results is thus based on an abstraction in which many arguably central aspects of the experimental vignettes are absent. In particular, two aspects should be highlighted.

First, subjects are *not* making their intentionality judgments on the basis of a description of a decision already made by a lone agent with certain beliefs and attitudes. For example, in the decision-point rendering we are presented with a description of an attitude and a motivating reason: “But the chairman doesn’t care at all about effect *x*. His sole reason for implementing the new program is that he believes it will make a lot of money for the company”. Test subjects, by contrast, read a statement made by the agent specifically in response to the interlocutor’s articulation of the side effect: “I don’t care at all about harming the environment. I just want to make as much profit as I can”. The decision-point rendering is an accurate rendering of the Chairman scenario only if the description of the attitude is equivalent to the corresponding statement in the vignette, such that what test subjects take from reading the chairman’s conversational contribution is a piece of information about his attitude toward the side effect.

Secondly, there are important differences between the sequential *Help* and *Harm* vignettes other than the moral valence of the side-effect. In particular, as we will argue, the characters perform very different social actions in the two vignettes. One salient difference is the introduction of the side effect into the conversation and the agent’s response to its suggested relevance. We discuss these differences in some detail in the “[Knobe’s Vignettes as Accounts of Action Sequences](#)” section. But first, we present the analytic perspective that informs our discussion.

## Analytical Approach: Ethnomethodology and Conversation Analysis

Our analyses are indebted to the analytic approach and empirical results of the two closely related fields of conversation analysis (Sacks et al. 1974; Sacks 1992; Schegloff 2007) and ethnomethodology (Garfinkel 1967). A central concern in ethnomethodology is how social actors in interaction achieve social order. A crucial resource for this achievement are the ways in which social actors render their actions intelligible and recognizable, or *accountable* (Garfinkel 1967). Applied to texts, such intelligibility has been described as “both a feature and an outcome of a set of practices inscribed and evident in, and relied upon, by the text” (Jayyusi 1991: 166). The interest of ethnomethodological analyses of texts lies in uncovering how “reading is made possible” and how the sense and intelligibility of a text is generated. Such analyses have challenged the view of texts as passive “conduits” (Watson 2009), casting doubt on the view of reading as an uncovering of abstract scenarios behind concrete textual details. Dorothy Smith’s (1990) notion of the *active text* is central here. This notion is meant to convey that any text has “its own structuring effect, that actively potentiates the sense of some phenomenon” (Watson 2009: 14). To illustrate, Lena Jayyusi (1993) examines a textual report of a shooting incident, and a courtroom hearing concerning a similar case, showing how narratives are built to suggest particular interpretations of intent and responsibility regarding actions and their consequences. In particular, the details of action descriptions are shown to be central to how the text “actively potentiates” certain kinds of intent ascription with regard to described actions.

This approach to the analysis of texts has consequences for how we understand the Knobe effect and discussions of it within experimental philosophy. The approach alerts us to the potentially problematic nature of the relation between the decision-point rendering and the sequential rendering in Knobe (2003); many textual details must be disregarded or considered irrelevant to treat the two renderings as equivalent. It also alerts us to important differences between the texts mediating the experimental conditions (*Harm and Help*), differences that go far beyond the simple substitution of ‘help’ for ‘harm’.

We use Conversation Analysis (CA) as a resource for our analyses because Knobe’s vignettes present a conversation between two or more fictional characters. This is also true of most vignettes used in later studies of the Knobe effect. While the test subjects are not themselves involved in a conversation, each of them is requested to judge the intentionality of a side effect that is the result of one.

CA has demonstrated how social actions are necessarily responsive to prior turns in a conversation and set up normative and interpretative constraints on the actions that follow. People engaged in conversation perceive and respond to actions taking these constraining properties into consideration; for example, any action following something that is recognizable as a question will be heard as an answer, unless due effort is taken in the design of the action to show how it is not an answer. Conversely, an utterance’s recognizability as an answer hinges on the conversationalists’ recognition of the prior turn as a question (and “answering” reflexively treats it as such). Emanuel Schegloff and Harvey Sacks put it as follows:

[A] pervasively relevant issue (for participants) about utterances in conversation is ‘why that now’, a question whose analysis may [...] also be relevant to finding what ‘that’ is. That is to say, some utterances may derive their character as actions entirely from placement considerations. For example, there do not seem to be criteria other than placement (i.e., sequential) ones that will sufficiently discriminate the status of an utterance as a statement, assertion, declarative, proposition, etc., from its status as an answer. Finding an utterance to be an answer, to be accomplishing answering, cannot be achieved by reference to phonological, syntactic, semantic, or logical features of the utterance itself, but only by consulting its sequential placement, e.g., its placement after a question. (Schegloff and Sacks 1973: 299)

Since Knobe’s (2003) analysis hinges on the ways in which test subjects understand the actions of a fictional agent, the sequential placement of the actions in question appears from this perspective to be critical not only to the test subjects’ assessment of the actions but also to their understanding of *what* those actions are in the first place. Identifying what action is being performed and assessing its consequences for the interaction are sequential matters. We assume that Knobe’s subjects will draw on their understanding of conversation in making intentionality judgments. This understanding makes them, for example, sensitive to normative constraints set up in a conversation by a *directive* or an *objection*, or to the interactional significance of words such as ‘but’ and ‘and’. On this view, subjects will read the actions depicted in the experimental vignettes with an eye to the question ‘why that now?’ and turn to sequential considerations for answers.

CA is replete with demonstrations of the interactional significance of seemingly ‘trivial’ details in the composition of individual turns. For instance, Heritage et al. (2007) demonstrate “the difference one word can make” in the language practices physicians use to inquire about patients’ possible concerns (after the official business of the visit has been concluded). “Do you have any more concerns” discourages the articulation of further problems, while “do you have some more concerns” encourages elaboration. The precise design of turns, here the use of “some” versus “any,” is thus consequential in interaction. Further illustrating the significance of details in CA, Schegloff describes “turn-initial components” as “little objects that do a piece of sequential work” (Schegloff 1987: 71), and Heritage et al. (2007) discusses how “sequential markers” of various kinds convey relations between current and prior contributions. For instance, whether an utterance is heard as an objection, a question or an encouragement may turn on such details as an inserted “but”. All this invites a reconsideration of what the relevant differences between the experimental conditions might be with regard to the sequencing and composition of talk and action in the vignettes.

Central to how intentional action is operationalized in Knobe’s study are the ways in which the characters in the vignettes express their attitudes and states of knowledge: their degree of “caring” and their “knowing” about the side effect. The agent’s beliefs and motivations are communicated to subjects as articulated by the agent at particular junctures in a sequentially unfolding narrative. Following a CA-perspective, this psychological aspect of the actions depicted in the vignettes can be analysed in terms of the social actors’ conversational use of affective and mental vocabularies.

This topic is addressed by the branch of interactional research (based on ethnomethodology and conversation analysis) known as Discursive Psychology (Edwards, 2006). Discursive psychologists argue that psychological terms are not only and not primarily used in the production of *reports* about the inner states to which particular mental terms refer, but rather as part of the rhetorical construal of events, actions and circumstances. When people say that they “do not care” for instance, this is usually not a neutral report about a state of mind, but an action doing something in the local sequential context. This is usually also how it will be interpreted. Interlocutors will make sense of the utterance in terms of what came before and what social action the utterance is heard as performing. For them, the statement raises the question of what locally relevant action or project the speaker is addressing—supporting, undermining, or altering—with the claim of ‘not caring’. As Derek Edwards puts it: “Rather than starting with the assumption that what people say is an outward *expression of* what they internally know, intend and think, [discursive psychology] examines the ways in which intentionality, states of mind, motives and thoughts (etc.) are matters *at stake in* discourse and social interaction” (2006: 46). How the characters of the vignettes “do things with words” (Austin 1975) is then a central question, and one to which we turn in the following analyses.

### Knobe’s Vignettes as Accounts of Action Sequences

In the following, we present the vignettes from Knobe (2003) in a way that makes the sequential development of the described actions and events explicit. The intended action and the side effect are presented as the final ‘turn’ of the conversation. Thereby we aim to make explicit the place of the outcome in the longer sequence portrayed in the vignette. We begin our analysis with the Chairman scenario, here presented with the turns of the conversation numbered. The *Harm* vignette is presented on the right, and the *Help* vignette on the left (VP = vice-president; CM = chairman).

<p>01 VP: We are thinking of starting a new program. It will help us increase profits <b>but</b> it will also <b>harm</b> the environment.</p>	<p>01 VP: We are thinking of starting a new program. It will help us increase profits <b>and</b> it will also <b>help</b> the environment.</p>
<p>02 CM: I don’t care at all about <b>harming</b> the environment. I just want to make as much profit as I can. Let’s start the new program.</p>	<p>02 CM: I don’t care at all about <b>helping</b> the environment. I just want to make as much profit as I can. Let’s start the new program.</p>
<p>03 ((They started the new program. Sure enough, the environment was <b>harmed</b>))</p>	<p>03 ((They started the new program. Sure enough, the environment was <b>helped</b>))</p>

To control for the specific features of the chairman case (e.g., subjects' attitudes towards environmental damage and corporations), Knobe (2003) also used another scenario, the Lieutenant scenario. The same results were obtained (LT=Lieutenant; ST=Sergeant).

<p>01 LT: Send your squad to the top of Thompson Hill!</p> <p>02 ST: <b>But</b> if I send my squad to the top of Thompson Hill, we'll be moving the men directly into the enemy's line of fire. Some of them will surely be killed!</p> <p>03 LT: Look, I know that they'll be in the line of fire, and I know that some of them will be killed. But I don't care at all about what happens to our soldiers. All I care about is taking control of Thompson Hill.</p> <p>04 ((The squad was sent to the top of Thompson Hill. As expected, the soldiers were moved into the enemy's line of fire, and some of them were killed.))</p>	<p>01 LT: Send your squad to the top of Thompson Hill!</p> <p>02 ST: If I send my squad to the top of Thompson Hill, we'll be taking the men out of the enemy's line of fire. They'll be rescued!</p> <p>03 LT: Look, I know that we'll be taking them out of the line of fire, and I know that some of them would have been killed otherwise. But I don't care at all about what happens to our soldiers. All I care about is taking control of Thompson Hill.</p> <p>04 ((The squad was sent to the top of Thompson Hill. As expected, the soldiers were taken out of the enemy's line of fire, and they thereby escaped getting killed))</p>
---	---

In presenting the Chairman case, Knobe claims that *Help* is “almost exactly the same [as *Harm*] except that the word ‘harm’ was replaced by ‘help’” (2003: 191). Much of the argument in Knobe (2003) hinges on this claim. As we have indicated, however, there are other salient differences that may be relevant to how subjects' responses should be interpreted. An initial difference at the level of wording can be seen in the substitution of “and” for “but”. In sequential terms, the “but” marks the side effect as a possible objection to the development, while the “and” marks the side effect as additional support for the development. This difference is highly salient in the Lieutenant scenario where the initial turn is a directive, an order. The sergeant's turn 02 is phrased as an objection in *Harm*, while this phrasing is absent in *Help*. Hence, at the point where the side effect has been presented, *Help* and *Harm* have developed into two quite different sequences. In *Help*, support for the prior directive has been presented. But in



*Harm*, an objection to the suggested action has been produced by the interlocutor's highlighting of the side effect. In the Lieutenant scenario, this is an objection to an order. In the Chairman scenario, it is a complicating consequence of the suggested action articulated by the agent's interlocutor.

Discussions of the Knobe effect standardly proceed from the assumption that the only difference in wording between the vignettes is the 'harm'/'help' difference (see, e.g., Adams and Steadman 2004: 174; Hindriks 2008: 632; Nadelhoffer 2010: 282; Sripada 2010: 161). For example, Cova first quotes the *Harm* vignette of the Chairman scenario and then states that other subjects were "given the same vignette, but this time with the word 'harm' changed into 'help'" (2016: 122). Of course, not mentioning other differences does not necessarily reflect an oversight. A more charitable reading is that commentators simply point out what is assumed to be the only relevant difference between the vignettes. It would be nit-picking, one might think, to highlight the additional difference between 'but' and 'and'. While we do not claim that this additional difference plays a decisive role in generating the response asymmetries, we want to suggest that it contributes to the narration of sequences of conversational actions in the vignettes.

An important aspect of the action of bringing up the side effect is that the interlocutor's own assessment of it is evident. The social action of bringing it into the conversation displays the interlocutor's stance toward the side effect—through markers such as "but" and "and"—as well as highlights it as relevant to the decision to be taken. In *Harm*, the characters treat it as a complicating factor for the intended action, but not in *Help*. Regardless of what presumed 'moral valence' would be assigned to the side effect by the subject reading the vignette, there is a proximal valence already assigned to it by the vignette's characters. In the decision-point rendering, this difference between *Harm* and *Help* is absent.

In the sequential rendering, after the turn in which the interlocutor articulates the possible relevance of the side effect, we have the agent's response. How we understand this response will turn out to be crucial, and much of the following discussion will therefore focus on various aspects of its design and its sequentially conditioned significance. To begin with, we suggest that, in *Harm*, subjects presumably read the agent's response as a counter to the previous turn. In the Lieutenant scenario, for instance, we have the utterance "Look, I know that [...], but I don't care at all...". When "look" is used as a turn-initial component of a response to a prior turn, it "marks a disjunction and a redirection of the talk" (Sidnell 2007: 387). Furthermore, as Marjorie Harness Goodwin (1982) has observed, responses such as "I know" and "I don't care" are used in disputes to argue for the irrelevance of the prior contribution. They are "especially apt moves following statements whose truth values are not an issue, such as warnings [and] criticisms" (85). A construction such as "I know" also works rhetorically as a riposte, undermining the status of the prior turn as "news" (Goodwin 1982). In the vignette, the agent is placed precisely in this position; he does not question the truth-value of the articulated side effect, and so produces the objection by other means. What is central is thus the social action of countering the side effect's presented status as a relevant objection, rather than any literal informational content of the uttered sentence.

The agent's objection to the relevance of the side effect, one could also add, builds on so-called *extreme case formulations* ("I don't care *at all* about...," "*all* I care

about is...”). In everyday speech, such formulations typically occur in argumentative sequences in which an “adversarial or defensive stance” is conveyed (Pomerantz 1986: 220). More generally, Edwards (2000) shows that, in disputes, participants typically deploy extreme case formulations to display that a locution is used non-literally.

It thus seems plausible that subjects understand the agent’s reply in *Harm* as implying that the agent is treating the interaction as an argument or potential dispute with two possible outcomes. In rejecting the practical relevance of the prior turn, the agent’s utterance thus becomes a potentially important causal factor in bringing about the negative side-effect. If no such rejection had been issued, the decision would perhaps have turned out differently. In conclusion, a crucial component in *Harm* is the action of rejecting a suggested complication to the intended action. The matter at stake that this action addresses is the ultimate decision as to whether or not to go ahead with the plan to send the squad to the hill or to start the new program.

In *Help*, there is no argument about whether or not to implement the action of moving the squad or starting the new program. The articulation of the side effect does not point to any complication associated with the suggested action, and the agent’s reply does not function as a consequential response to the practical implication of the prior turn (that is, that they should move forward with the decision). The agent’s response, however, still signals disagreement. How can we account for this objection to an expression of support? If the argument is not about the decision under consideration, then what is it about?

We noted above that the lieutenant’s reply (“Look, I know...” and so on) serves to reject the relevance of the prior contribution. What significance could such a rejection have in a situation where the decision itself is not at stake? Note that at the point where the side effect has been articulated in *Help*, it is unclear what the further intent or purpose of the suggested action is. To speak of a main intended action and a side effect is premature at this point in the narrative. In the chairman version, the vice president presents *two* effects of the new program, without any indication whatsoever of their relative significance. In fact, the vice president is easily heard as presenting profit and environmental improvement as the two reasons for starting the program. Note that the two consequences are appended to the articulation of the plan, thus functioning as an *account* supporting the plan. In the lieutenant scenario, the sergeant’s exclamation that lives will be saved can even be heard as a candidate formulation of an unarticulated further intent or purpose behind the order, that is, that moving the squad to Thompson hill is done in order to save the lives of the soldiers. It is only apparent in the sergeant’s subsequent reply that the squad is moved in order to “take control” of the hill, a purely military-strategic purpose as opposed to a safety-related one. Similarly, in the Chairman scenario, it is only with the chairman’s reply to the vice president’s initial turn that we get any sense that one of the articulated consequences should be considered an unimportant “side effect” rather than one reason for engaging in the intended action. In *Harm* on the other hand, the two consequences are differentiated by the *x but y* construction, conveying that *x* is *the* motivation for the action, permitting a characterisation of the other as a “side effect”.

Implicit in the interlocutor’s turn in the *Help* vignette of the Lieutenant scenario, in which the side effect is mentioned, is thus a suggestion of what the action’s further intent or purpose is. This suggestion is then *corrected* by the agent’s reply. The

interlocutor's candidate interpretation of the prior directive's motive is rejected, followed by the articulation of a different motivation. In the Chairman scenario, where the side-effect is mentioned as a suggested reason in favour of starting the new program, the ambiguity inherent in the vice president's presentation of the program is resolved by the Chairman's rejection of the relevance of the environmental effect, followed by a repetition and reinforcement of the other purpose mentioned for starting the new program, namely making profit.

This analysis shows that the interlocutors are *not* debating the particular decision to be taken in *Help*. They are rather negotiating, and the agent corrects and clarifies *what* the relevant description of the intended action is, that is, what its further purpose is. In the Lieutenant scenario, are they 'saving [their] men' or 'taking control of Thompson hill'? In the Chairman scenario, are they starting a new profitable program or are they engaging in good business and environmental work? The agent's reply in the vignette serves to disambiguate what should be regarded as the main, in fact the only, purpose of the endeavour: taking control of the hill, making profit. *Help* thus contains, as an emergent feature of the narrated sequence, a kind of participant perspective on how the relative intentional status of the two articulated consequences should be evaluated. By resolving an initial ambiguity, the sequence itself plays out as a *clarification* of what the agent takes to be "intentional".

In conclusion, a central component of *Help* is the correction of a suggested further intent or purpose of an intended action (the Lieutenant scenario) or the rejection of a suggested reason favouring an action (the Chairman scenario). The correction or rejection clarifies for relevant parties what should be treated as the accountably relevant intentions or reasons for action, and it resolves ambiguities regarding intent. The matter at stake addressed by the correction or rejection is thus the vignette characters' shared understanding of intent itself. This plausibly explains why the beneficial side effect is not judged to be intentional. The sequences in the *Help* vignettes suggest a certain interpretation of "intentionally". This is an interpretation according to which intentional action simply means an action that is intended. Given this meaning of intentional, side effects are not intentional.

We hope to have made clear that, apart from the differences in wording, *Harm* and *Help* represent very different action sequences. In reducing the vignettes to a representation of an event involving a single agent, a single intended action, a set of attitudes and a side-effect, Knobe and others gloss over differences that are potentially important for subjects' intentionality judgements. Here, we would like to highlight one salient difference in particular: In *Help*, the conversation depicted in the vignette is about intentionality in the sense of intent or purpose.<sup>3</sup> The story revolves around the ways in which the intent of an agent should be properly described. This story is told via the characters' suggesting purposes or reasons for

<sup>3</sup> Regarding the *Help* vignette of the Chairman scenario, none of the subjects judged that it was the Chairman's "intention" to help the environment. Twenty percent of the test subjects judged that the Chairman helped it "intentionally" (Knobe 2004). Regarding the *Harm* vignette, 29% of the test subjects judged that it was the Chairman's intention to harm the environment, and 87% judged that the Chairman harmed it intentionally.

action which are then either validated or corrected/rejected. *Harm*, on the other hand, is not a story about intent. The vignette's characters do not treat the agent's intent as a topic of discussion; it is rather treated as a given. *Harm* is primarily a story in which the agent is placed in a position of *choice*, where reasons favouring both options are brought forward by the agent's interlocutor: aborting an intended action when a particular consequence of the action is presented as relevant or moving forward in spite of it.

*Harm* is thus a story about choice and *Help* is a story about further intent. When subjects are asked to assess whether or not the so-called "side effect" was brought about "intentionally," this question is thus posed in two very different contexts. Given the difference in context, one could ask whether subjects make sense of the question in the same way in the two experimental conditions. Our suggestion is that "intentionally" means something different in the context of choice compared to a context where the clarification of the agent's further intent is the main issue.

To argue this, we need to deal with variations of Knobe's study that show that subjects do not always rate side effects as intentional in stories focusing on choice. In studies using vignettes where the agent expresses regret at the prospect of bringing the side effect about, the judgment of intentionality that would be assigned to the negative side effect in otherwise similar *Harm* vignettes is moderated (e.g., Sverdlik 2004; Cova 2014; Knobe and Kelly 2009). For instance, Knobe (2007) found no difference in intentionality ratings using such a vignette when compared to the original *Help* vignette. Let us look briefly at one such scenario [adapted from Mele and Cushman (2007)].

(AL = Al; AN = Ann):

```

01 AL: You know, if you fill in that
      pond in the empty lot next to
      your house, you're going to
      make the kids who look for
      frogs there sad
02 AN: I know that I'll make those
      kids sad. I like those kids,
      and I'll definitely regret
      making them sad. But the pond
      is a breeding ground for
      mosquitoes; and because I own
      the lot, I am responsible for
      it. It must be filled in.
03 ((Ann filled in the pond, and
      sure enough, the kids were
      sad.))

```

Did Ann intentionally make the kids sad? MVH would predict that she did, since the side effect is ‘morally bad’ from the standpoint of the typical subject. As it turns out, however, subjects respond in a way similar to the original *Help* scenario; a majority judges the side effect to be unintentional. How come? One common way of interpreting this result is that it shows how the agent’s attitudes towards the side effect influence intentionality ratings (see Cova 2016). This is usually expressed by saying that an element of *regret* is introduced. Similar to discussions of the original *Help* and *Harm* scenarios, the agents’ utterances in a conversational sequence are thus considered primarily as vehicles of informational content regarding the agents’ mental states.

What such a treatment does not acknowledge is that Ann’s expressed attitude is a component of her *justification* of the decision, which is very different from the agent’s response in the original *Harm* scenarios. Ann, in turn 02, does not flatly *reject* the objection. Instead, she produces a complex reply that strikes a balance between acknowledging the objection and retaining the (implicitly) intended action of filling the pond. She appeals to circumstances that favour that she fills the pond: the presence of mosquitos, her ownership of the pond, and the obligations that come with ownership. The result is a portrayal of the action of filling the pond as an action that she is forced to take (“It must be filled in”). She thus presents a situation where there is actually no choice between two live options, at least not insofar as she is to remain a rational moral agent. As part of a justification, the regret component is a rhetorical device by which the agent demonstrates the unavoidable status of the action, while acknowledging and accepting the prior speaker’s assessment of the side effect. Of course, the agent might be taken to sincerely communicate an affective stance—an expression of sadness and regret at the prospect of bringing the side effect about. However, what is important is the way in which the expression helps the agent justify her action and portray the situation as one of constrained choice.

Hence, results from studies of ‘regret’ arguably provide additional support for our suggestion that ‘intentionally’ has a different meaning in the context of choice than in a context where the main issue is the clarification of the agent’s further intent or reason for action. Viewing bad side-effect vignettes (*Harm* and *Regret*) as revolving around choice, and consequently interpreting ‘intentionally’ as connected with moral responsibility, one would expect justifications and accounts to matter for how subjects understand and assess the central actions in the vignette (a point to which we will return). In this, we are in line with Mele and Cushman (2007) in their acknowledgment that subjects may judge the agent to be justified in her actions and moderate intentionality ratings accordingly, rather than in response to the agent’s expressed attitudes as such.

## Lone Agent Scenarios

Some vignettes used to study the Knobe effect do not describe a conversation between the agent and an interlocutor. Instead, they involve a lone agent making and acting on a decision. These vignettes nevertheless diverge from a decision-point rendering in important ways. What is described is a deliberative process rather than

a set of attitudes and a momentary decision (for examples, see some of the vignettes in Nadelhoffer 2006). The different events and actions leading up to the decision are presented as sequentially developed in the course of the agent's practical reasoning about what to do. Like Knobe's vignettes, their temporal character and the sequence of actions depicted put them at odds with a decision-point rendering.

As in Knobe's vignettes, the effect of selectively highlighting a particular side effect undermines claims about how subjects are evaluating the valence of a side-effect also in lone agent scenarios. Even when no conversation is presented, merely stating that the agent does not care at all about a side-effect  $x$  is unlikely to make the subject interpret the scenario such that the agent takes  $x$  to be completely irrelevant. Consider the following vignette, which was used by Thomas Nadelhoffer to find out whether or not subjects would judge that a neutral side-effect was brought about intentionally (all emphases except 'definitely' are added by us):

A sniper has been ordered to kill an enemy commander. So, after getting himself into position, he finally has the enemy commander in his sights. Before he pulls the trigger, *however*, the sniper *realizes* that the gunfire will *definitely* cause the barrel of his gun to get hot. But the sniper doesn't care at all whether the barrel of the gun is hot, *he doesn't have to touch it anyway*. So, he pulls the trigger – thereby shooting and killing the commander. And, as the sniper expected, firing the gun caused the barrel to heat up. (Nadelhoffer 2006: 149f.)

Sixty-eight percent of Nadelhoffer's subjects judged that the sniper intentionally caused the barrel to heat up. Nadelhoffer takes this result to throw doubt on the view that a foreseen side-effect is only intentionally brought about if, other things being equal, the agent has a reason not to bring it about. We do not think this result shows this at all. While the vignette explicitly states that the agent "doesn't care at all" about the side-effect, it nevertheless describes a process of deliberation in which the sniper first considers the heating of the barrel to be a bad thing (other things being equal). After all, if the prospect of a heated barrel is not at all relevant to the sniper's activity, then it would not make sense to state that he "realizes" that the barrel will be heated. Only in light of the further realisation that he will not have to touch the barrel does he conclude that he need not worry about it. In fact, merely mentioning the side-effect of the barrel's becoming heated signals to the subject that it has an important place in the sniper's deliberation. In line with our analyses of *Harm* and *Help*, we should pay attention to the sequential development of the narrative. Once we do this, it is clear that there is a point in the narrative where the agent is undecided about what practical significance should be assigned to the 'side effect'. Note also that the realization of the side effect is introduced by "however," marking the realization as in some way conflicting with the intended action (parallel to the "but" in Knobe's original *Harm* vignette). We should therefore expect that a reader would take the heating of the barrel to be a side-effect that is prima facie relevant to the sniper's decision and future course of action.

Even in lone agent scenarios, the vignettes typically present the process of deliberation in a conversational format, as an "inner" dialogue. In this inner dialogue, the agent says things to him- or herself and thereby comes to realize that some consideration or other is relevant to her decision (see Knobe and Kelly 2009; Nadelhoffer

2006). For example, consider the following *Terrorist* vignette from Knobe and Kelly (2009):

A terrorist discovers that someone has planted a bomb in a nightclub. There are lots of Americans in the nightclub who will be injured or killed if the bomb goes off. The terrorist says to himself, “Whoever planted that bomb in the nightclub did a good thing. Americans are evil! The world will be a better place when more of them are injured or dead.”

Later, the terrorist discovers that his only son, whom he loves dearly, is in the nightclub as well. If the bomb goes off, his son will certainly be injured or killed. The terrorist then says to himself, “The only way I can save my son is to defuse the bomb. But if I defuse the bomb, I’ll be saving those evil Americans as well... What should I do?”

After carefully considering the matter, he thinks to himself, “I know it is wrong to save Americans, but I can’t rescue my son without saving those Americans as well. I guess I’ll just have to defuse the bomb.”

He defuses the bomb, and all of the Americans are saved. (Knobe and Kelly 2009: 177)

The first paragraph frames the story by presenting the terrorist’s disdain for Americans and the positive moral valence that he assigns to the prospect of their being killed. In the second paragraph, we find what corresponds to the initial exchange between the chairman and the vice president in which the intended action (the defusing of the bomb in order to save the son) and the foreseen side effect (that the Americans are saved) are presented. As in the original vignettes, the side effect has a clear practical/moral valence that is internal to the story itself (ascribed by the agent). But unlike in the original vignettes, this internal valence is unlikely to be aligned with the valence that the subject assigns to it.

The third paragraph describes the agent’s reasoning in reaching his decision. Here, again, there is a salient difference compared to the original *Harm* vignettes. The fact that the side effect will be brought about if the terrorist defuses the bomb is presented as an objection that the terrorist himself considers in his practical deliberation. However, unlike in the original *Harm* vignettes, the terrorist does not go on to reject the practical relevance of this. Instead he explicitly acknowledges its relevance by thinking to himself, “I know it’s wrong but...I guess I’ll just have to defuse the bomb”. After all, the side effect goes against his previously expressed attitude towards the impending death of the Americans in the club. But since the terrorist’s son is in the nightclub, he must nevertheless defuse the bomb. Knobe and Kelly (2009) found that people tended to judge that the terrorist did *not* save the Americans intentionally, and take this to support the idea that it is the moral valence of the side effect assigned by the subjects that explains the asymmetry of responses in the original *Help* and *Harm* vignettes. Recall, however, that justifications and expressed regret moderate intentionality ratings in vignettes with “bad” side effects. Rather than the reversal of moral valence, the terrorist’s acknowledgement of the badness of the side-effect and the justifications he provides can explain the result.

## Discussion

We will now turn to a discussion of our analyses. We start by turning briefly to a set of issues addressed in experimental philosophy concerning how the Knobe effect should be characterised and explained.

### Explaining Response Asymmetries: Actions and Evaluative Perspectives

Knobe's original characterisation of the observed response asymmetries was that "people seem considerably more willing to say that a side-effect [of an agent's intended action] was brought about intentionally when they regard that side-effect as bad than when they regard it as good". Note that the difference between the experimental conditions is caught by the brief phrases "they regard [the side effect] as bad" and "they regard it as good". Our analysis suggests a somewhat more cautious and less general formulation of the experimental results. On the one hand, people tend to judge a side-effect as intentional if it has first been brought up as a consideration against a course of action, but then rejected by the agent as irrelevant in a turn that instead stresses the exclusive importance of an intended effect. As we have seen, this tendency is absent when the agent acknowledges the side-effect as relevant to her decision but justifies why she nevertheless has to bring it about. On the other hand, when the agent instead rejects a particular consequence as irrelevant after it has been brought up as a candidate further purpose of the intended action or as a reason for performing it, subjects tend not to judge that it has been brought about intentionally. This characterisation of the results is not as economical as Knobe's original, but we believe it does better justice to the salient differences between the *Harm*, *Regret*, and *Help* vignettes.

Our characterisation also provides a more nuanced picture of the "stimulus" that test subjects are exposed to and of the differences between experimental treatments that are plausibly relevant to the differing intentionality judgments. Our argument springs from the observation that the vignettes contain *sequences* of actions, rather than a single action. Scaife and Webber (2013) offer an account that rightly emphasizes the role of the actions that are performed by the agent during deliberation. According to their "consideration hypothesis," what explains the Knobe effect is whether or not the chairman takes the side-effect into consideration when making his or her decision to start the development: The side effect is intentionally brought about if the agent "considers" it in the deliberation leading up to the agent's choice and taking something into consideration is an intentional action. According to Scaife and Webber, the chairman considers the side-effect in the *Harm* vignette and therefore brings it about intentionally, whereas it is not clear whether he considers it in the *Help* scenario. They argue that, in *Help*, the fact that the side effect speaks in favour of the intended action means that subjects are more likely to take the chairman's claim to "not care at all" literally, leading the subjects to conclude that the side effect has not been considered by the chairman.

However, it is clear from our analyses that "considering" is too broad a category to adequately account for what actually happens in the vignettes. Claiming that



the chairman *considers* the side effect in the *Harm* vignette distorts what happens. While the chairman does register that the vice-president has spoken and offered an objection, he actively *rejects* this objection, undermining the alleged relevance of the side effect for the deliberative process. He does so, moreover, by appealing to his own indifference. Contrast this, for instance, with the way in which the terrorist considers the side-effect that Americans will be saved in the *Terrorist* vignette, where the side effect is acknowledged as relevant by the terrorist, who brings it in as a relevant consideration in the process of deliberation. When the terrorist nevertheless decides to defuse the bomb, this is done after a stretch of reasoning in which the decision is *justified* by the agent (as the only way to save his son) and accompanied by an expression of regret regarding the foreseen side effect. This response is a very different type of action from the chairman's rejection. In light of this, it is unsurprising that Scaife and Webber's consideration hypothesis fails to account precisely for "regret" cases such as the terrorist vignette (see Cova 2014).

Other authors have noted that, in addition to deciding "to implement a new program" (Knobe 2003; the decision-point rendering), the chairman also performs another action in the *Harm* vignette: He intentionally ignores a normative reason against the action, or flouts a moral or social norm (e.g., Hindriks 2014; Holton 2010; Nanay 2010; Scaife and Webber 2013). A central issue here is which perspective of moral evaluation is crucial for creating the asymmetry in responses. According to MVH, it is the subject's evaluation of the side-effect that matters but, according to other hypotheses, it is the subject's understanding of the *agent's* evaluation of the side-effect (Turner 2004). Our argument here is clearly in favour of paying close attention to the agent's perspective. The agent does not simply ignore or flout an externally applied social norm, since normativity and evaluation are manifest features of the vignette itself. The subject's evaluative perspective is thus rendered obsolete. But neither should we consider the conversational actions of the agent simply as reports on his or her attitudes and evaluative perspectives. Rather, the agent responds to an argument in which the side effect is suggested to be somehow relevant in relation to the intended action. In *Harm* the positioning of the side effect as a possible complication engenders an argumentative decision-making sequence, and a focus on choice. In *Help* it is instead the clarification of intent that is in focus. In none of these contexts, as we have argued, are attitudes and moral standpoints central; instead we have shown how the vignettes can be described in terms closer to the surface of the narrated events, without necessitating considerations of external or underlying evaluative perspectives.

## Everyday Reasoning and Experimental Vignettes

In analysing Knobe's vignettes, we drew on some results from discursive psychology. But discursive psychology might also be relevant for understanding Knobe's experimental study itself. Discursive psychologists ask how people explicitly address the intentional status of actions in various circumstances. How are "intend," "intentional" and other mental terms used in everyday speech? In ordinary interactions, intent is typically topicalized in situations where participants address

*problematic* features of an action, and where responsibility and blame are assigned (Jayyusi 1993). This concern with intent in the context of negatively evaluated actions can be further illustrated by studies of legal settings. Edwards describes a set of topicalizations of intent in police interrogations as “descriptions of actions and their intentionality, where what is intended, foreseeable, known, accidental or incidental, is defeasibly assembled *in and for the performance of police work*, and with regard to the action categories of relevant law” (2008: 45, emphasis in original). Now, the task assigned to the test subjects in Knobe’s study raises the issue of intentionality explicitly. They are asked to assess whether or not a particular described consequence of an action was intentionally brought about. What might this question reasonably mean in the context of a particular vignette?

The two original experimental conditions, *Harm* and *Help*, provide different contexts for the ways in which the subjects interpret their task; what the question about intentionality is taken to mean in the two contexts might then differ significantly. We have argued that the “side effect” plays very different roles in *Help* and *Harm*. In the *Help* vignettes, the side effect is a candidate further purpose or reason for action whose status as such is explicitly corrected or rejected. In the *Harm* vignettes, it is a negative consequence that the agent in brief argumentation with his interlocutor rejects as irrelevant to his decision. *Help* is a story about intent, while *Harm* is a story about choice. We have argued that the *Help* and *Harm* vignettes evoke different notions of ‘intentional’.

We have seen how intentionality ratings in variations of *Harm* are sensitive to the details of the agent’s response to the articulation of the side effect—acknowledgments of bad side-effects and justifications, for instance, moderate the ascription of intentionality. The question seems, in *these* scenarios, to be interpreted in a way that connects intentionality with moral responsibility and accountability. Adams and Steadman (2004) observe that the word intentionally has such a usage, which is also in line with empirical studies of topicalisations of intentionality (Edwards 2008). Indeed, subjects’ responses in *Harm* have from the outset been shown to correlate with assignments of blame (Knobe 2003). More recently, Hindriks et al. found that “the degree of responsibility [blameworthiness] ascribed correlates with the chance with which intentionality is attributed in the harm condition, but not in the help condition” (2016: 211). We take this to reflect the fact that “intentionally” is interpreted differently in the contexts provided by *Help* and *Harm*.

The story in *Harm* is precisely about choice and responsibility. The nature of the action accounts provided by the characters in the vignette should matter for intentionality ratings in such a context. We have argued that the turn in which the agent responds to the presentation of the side effect is critical. We can now see that this is because the agent’s response is the place in the narrative where an *account* can be provided. In the *Regret* scenario, the agent appeals to reasons, contextual circumstances and obligations beyond the agent’s immediate control (Mele and Cushman 2007). This form of justification appears to have a measurable effect on subjects’ intentionality ratings compared to *Harm*, where the agent’s response hardly qualifies as an account at all. It is widely accepted that for an agent to be morally responsible for an action or outcome it is normally sufficient that he or she has the relevant kind of control over, and knowledge of, the outcome in question. If the action or outcome

is bad in some way and the agent lacks relevant excuses, then he or she is also blameworthy for it. We take it that there is a sense of ‘intentional’ that is relevant in contexts where agents bring about consequences through choices for which they are both responsible and blameworthy (see also Duff 1982; Hindriks 2008; 2014).

Subjects’ sensitivity to accounts in the context provided by vignettes such as *Harm* and *Regret* can be further illuminated by considering the case of intent ascription in criminal law. This is a setting where the establishment of *mens rea* is critical, and where intentionality is treated as “a graded property from full premeditation to ‘recklessness’ with regard to consequences” (Edwards 2008: 183). Rather than only being concerned with what was intended in a narrow sense, there is an orientation to assessing the individuals’ responsibility for an action’s consequences. An extract from a police interrogation illustrates this. After having obtained a set of answers from a suspect regarding the act of punching and kicking a car, in which the suspect claims to have kicked the car to “take [his] temper out” (Edwards 2008: 187), the police interrogator says: “So you’re tellin’ me you kicked the car door. Um... you didn’t intend to cause damage: But—you’re aware that—by kickin’ something you can dent it” (187f.). By partialling out the act, the intention, and mental states such as awareness of consequences, and securing a confirmation from the suspect, the basis for determining culpability is provided. The damage was perhaps not intended but the act was accompanied by an awareness that the damage was a likely consequence. In such cases, as we have argued in relation to *Harm*, the accounts provided by the agent are crucial to how participants understand the agent’s responsibility for the consequence. This is what we take subjects to be concerned with in addressing the *Harm* vignette; they are weighing in awareness to determine culpability (see Hindriks et al. 2016).

## Experimental Philosophy and Ethnomethodology

We have tried to show that *Help* and *Harm* create different contexts for the application of questions of intentionality, an argument that complicates the prospect of finding a unified theory of “the folk concept of intentional action” that can account for both. To recall, Adams and Steadman question whether Knobe’s “ordinary language surveys of folk judgments have accessed core concepts of intentional action” (2004: 173). The question is whether experimental vignettes that provide such access can be constructed at all. It seems likely that, regardless of how vignettes are designed, issues such as the ones we have raised will crop up in *some* way. The mere fact that the vignettes are written in natural language arguably complicates the construction of variables that can be manipulated without simultaneously altering holistic aspects of the vignettes so that the experimental conditions come to represent scenarios too divergent to be reliably compared (as in *Harm* and *Help*).

A failure to access “the core notions of the cognitive machinery that underlies intentional action” (Adams and Steadman 2004: 174) need not be something lamentable. The varied and seemingly conflicting uses of intentional terms in ordinary interactions can be a topic of study in its own right. In addition to investigating ‘core concepts’ experimentally, philosophers should consider “the rich surface of

social interaction” (Edwards 2006) as a source of empirical data directly. From an ethnomethodological perspective, exploring that rich surface does not imply a loss in depth. Indeed, it is the central methodological recommendation. Edwards suggests that, “rather than asking is there a way of seeing below the surface to motives, thoughts and experiences, we can ask: are there procedures that participants have, for dealing with those notions?” (2006: 43).

Experimental philosophy, as described by Knobe and Nichols, aims at understanding what human beings are really like, explicitly recognizing “that such an inquiry will involve us in the study of phenomena that are messy, contingent, and highly variable across times and places” (2008: 3), and stressing the need for systematic empirical studies. At this point, one could imagine the relevance of a number of methodologies. Knobe and Nichols opt for “experimental investigations of the psychological processes underlying people’s intuitions about central philosophical issues” (2008: 3). However, an exclusive focus on experimental methods does not follow from the basic motivation of experimental philosophy. Throughout the present discussion, we have indicated several connections between the concerns of experimental philosophers and studies of ordinary talk-in-interaction, and we believe such connections could be further explored.

**Acknowledgements** We would like to thank two anonymous reviewers for Human Studies, as well as Ezio Di Nucci, Thor Grünbaum, Jon Hindmarsh, Frank Hindriks, Eric Laurier and John Michael for valuable input and thoughtful readings of earlier drafts. In addition, we would like to acknowledge the input from two reviewers for Philosophical Psychology who kindly reviewed an earlier version of the article. Olle Blomberg’s research was funded by project grant 421–2014–1025 from Vetenskapsrådet (the Swedish Research Council). Open access funding provided by Stockholm University.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Adams, F. (1986). Intention and intentional action: The simple view. *Mind and Language*, 1(4), 281–301.
- Adams, F., & Steadman, A. (2004). Intentional action in ordinary language: Core concept or pragmatic understanding? *Analysis*, 64(2), 173–181.
- Austin, J. L. (1975). *How to do things with words*. Oxford: Oxford University Press.
- Cova, F. (2014). Unconsidered intentional actions: An assessment of Scaife and Webber’s ‘Consideration Hypothesis’. *Journal of Moral Philosophy*, 11, 57–79.
- Cova, F. (2016). The folk concept of intentional action: Empirical approaches. In J. Sytsma & W. Buckwalter (Eds.), *Blackwell companion to experimental philosophy* (pp. 121–141). London: Wiley-Blackwell.
- Duff, A. (1982). Intention, responsibility and double effect. *The Philosophical Quarterly*, 32(126), 1–16.
- Edwards, D. (2000). Extreme case formulations: Softeners, investment, and doing nonliteral. *Research on Language and Social Interaction*, 33(4), 347–373.
- Edwards, D. (2006). Discourse, cognition and social practices: The rich surface of language and social interaction. *Discourse Studies*, 8(1), 41–49.
- Edwards, D. (2008). Intentionality and mens rea in police interrogations: The production of actions as crimes. *Intercultural Pragmatics*, 5(2), 177–199.
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.

- Goodwin, M. H. (1982). Processes of dispute management among urban black children. *American Ethnologist*, 9(1), 76–96.
- Heritage, J., Robinson, J. D., Elliott, M. N., Beckett, M., & Wilkes, M. (2007). Reducing patients' unmet concerns in primary care: the difference one word can make. *Journal of General Internal Medicine*, 22(10), 1429–1433.
- Hindriks, F. (2008). Intentional action and the praise-blame asymmetry. *Philosophical Quarterly*, 58(233), 630–641.
- Hindriks, F. (2014). Normativity in action: How to explain the Knobe effect and its relatives. *Mind and Language*, 29(1), 51–72.
- Hindriks, F., Douven, I., & Singmann, H. (2016). A new angle on the Knobe effect: Intentionality correlates with blame, not with praise. *Mind and Language*, 31(2), 204–220.
- Holton, R. (2010). Norms and the Knobe effect. *Analysis*, 70(3), 417–424.
- Jayyusi, L. (1991). The equivocal text and the objective world: An ethnomethodological analysis of a news report. *Continuum*, 5(1), 166–190.
- Jayyusi, L. (1993). Premeditation and happenstance: The social construction of intention, action and knowledge. *Human Studies*, 16, 435–454.
- Knobe, J. (2003). Intentional action and side-effects in ordinary language. *Analysis*, 63, 190–193.
- Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, 33(4), 315–365.
- Knobe, J., & Kelly, S. D. (2009). Can one act for a reason without acting intentionally? In C. Sandis (Ed.), *New essays on the explanation of action* (pp. 169–183). New York: Palgrave.
- Knobe, J., & Nichols, S. (2008). An experimental philosophy manifesto. In J. Knobe & S. Nichols (Eds.), *Experimental philosophy* (pp. 3–14). Oxford: Oxford University Press.
- Mele, A. R., & Cushman, F. (2007). Intentional action, folk judgments, and stories: Sorting things out. *Midwest Studies in Philosophy*, 31(1), 184–201.
- Mizumoto, M. (2018). A simple linguistic approach to the Knobe effect, or the Knobe effect without any vignette. *Philosophical Studies*, 175(7), 1613–1630.
- Nadelhoffer, T. (2006). Desire, foresight, intentions, and intentional actions: Probing folk intuitions. *Journal of Cognition and Culture*, 6(1–2), 133–157.
- Nadelhoffer, T. (2010). The causal theory of action and the still puzzling Knobe effect. In J. H. Aguilar & A. A. Buckareff (Eds.), *Causing human actions: New perspectives on the causal theory of action* (pp. 277–295). Cambridge: MIT Press.
- Nanay, B. (2010). Morality or modality? What does the attribution of intentionality depend on? *Canadian Journal of Philosophy*, 40(1), 25–40.
- Pomerantz, A. (1986). Extreme case formulations: A way of legitimizing claims. *Human Studies*, 9(2–3), 219–229.
- Sacks, H. (1992). In *Lectures on conversation* (Vol. 1 & 2). Edited by G. Jefferson. Oxford: Blackwell.
- Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50(4), 696–735.
- Scaife, R., & Webber, J. (2013). Intentional side effects of action. *Journal of Moral Philosophy*, 10, 179–203.
- Schegloff, E. A. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organization. In G. Button & J. R. E. Lee (Eds.), *Talk and Social Organization* (pp. 70–85). Clevedon: Multilingual Matters.
- Schegloff, E. (2007). *Sequence organization in interaction*. Cambridge: Cambridge University Press.
- Schegloff, E., & Sacks, H. (1973). Opening up closings. *Semiotica*, 8(4), 289–387.
- Sidnell, J. (2007). 'Look'-prefaced turns in first and second position: Launching, interceding and redirecting action. *Discourse Studies*, 9(3), 387–408.
- Smith, D. (1990). The active text: A textual analysis of the social relations of public textual discourse. *Texts, facts, and femininity: Exploring the relations of ruling* (pp. 120–158). London: Routledge.
- Sripada, C. S. (2010). The Deep Self Model and asymmetries in folk judgments about intentional action. *Philosophical Studies*, 151(2), 159–176.
- Sverdlik, S. (2004). Intentionality and judgment in commonsense thought about action. *Journal of Theoretical and Philosophical Psychology*, 24(2), 224–236.
- Turner, J. (2004). Folk intuitions, asymmetry, and intentional side effects. *Journal of Theoretical and Philosophical Psychology*, 24(2), 214–219.
- Utikal, V., & Fischbacher, U. (2014). Attribution of externalities: An economic approach to the Knobe effect. *Economics and Philosophy*, 30(2), 215–240.

---

Watson, R. (2009). *Analysing practical and professional texts: A naturalistic approach*. Farnham, UK: Ashgate.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.