Intentional action in ordinary language: core concept or pragmatic understanding?

Fred Adams & Annie Steadman

1. Introduction

Among philosophers, there are at least two prevalent views about the core concept of intentional action. View I (Adams 1986, 1997; McCann 1986) holds that an agent S intentionally does an action A *only if* S intends to do A. View II (Bratman 1987; Harman 1976; and Mele 1992) 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. Joshua Knobe (2003a) presents intriguing data that may be taken to support the second view.[1] Knobe's data show an asymmetry in folk judgements. People are more inclined to judge that S did A intentionally, even when not intended, if A was perceived as causing a harm (e.g. harming the environment). There is an asymmetry because people are not inclined to see S's action as intentional, when not intended, if A is perceived as causing a benefit (e.g. helping the environment).

In this paper we will discuss Knobe's results in detail. We will raise the question of whether his ordinary language surveys of folk judgments have accessed core concepts of intentional action. We suspect that instead Knobe's surveys are tapping into pragmatic aspects of intentional language and its role in moral praise and blame. We will suggest alternative surveys that we plan to conduct to get at this difference, and we will attempt to explain the pragmatic usage of intentional language.

We suspect that folk notions of intentional action are not clearly articulated. There are many factors required for an action to be performed intentionally. One of them involves the causal relation between an intention and the intended action. Not many folk would have very clear notions of counterfactual causal dependency of action upon intention necessary for intentional action on either view above. If an intention is connected by causal deviance to its conditions of satisfaction, the action is not done intention-

[1] In another article Knobe (2003b) makes the stronger claim that his experimental results do indeed show that moral considerations enter into the core concept of whether or not an action is intentional. If true, his results would tend to support View II. In the paper under discussion Knobe (2003a) only claims that there is an asymmetry in the folk concept and therefore it would be a 'mistake to ask for a general answer' (191).

Analysis 64.2, April 2004, pp. 173–81.

ally. Few folk would have clear notions of the exact relations of dependency between action and intention to block such causal deviance. Indeed, the exact relation of dependency is still in dispute among philosophers and cognitive scientists. However, almost everyone knows clearly that bad acts done intentionally are morally worse than bad acts done unintentionally. And almost everyone knows that saying 'you did that on purpose' is a social way to assign blame and of discouraging actions that one disapproves of. Hence, it is very likely that folk concepts of the pragmatic dimension of intentional talk are more richly understood than the core notions of the cognitive machinery that underlies intentional action.

2. *Knobe's experimental data*

In Knobe's first experiment, he handed out surveys to 78 people spending time in a Manhattan public park. Subjects were randomly assigned to one of two conditions: a 'harm' condition or a 'help' condition. Subjects read vignettes about actions that differed only by whether an actor helped or harmed the environment. The exact harm vignette was as follows:

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

In the 'help' vignette, Knobe gave the same scenario, replacing the word 'harm' with the word 'help' ('helping'). In both the 'help' and 'harm' conditions, subjects were asked to rate the amount of blame (or praise, respectively) the chairman deserved for harming or helping the environment on a scale from 0 to 6 and to say whether the chairman *intentionally* harmed or helped the environment.

The two conditions 'elicited … radically different patterns of responses' (192). In the harm condition, 82% of the subjects said the chairman intentionally harmed the environment. In the help condition, 77% said the chairman did not intentionally help the environment. The difference was highly statistically significant and stunning. Why the asymmetry? In the harm condition the folk judgments seem to accord with View II from above, while in the help condition the folk judgments seem to accord with View I. Why the difference?

Before we attempt to respond, we should point out that Knobe conducted a second experiment in order to validate his results. When Knobe

explains why he ran the second experiment, he says 'Perhaps the results obtained in experiment 1 can be explained in terms of some highly specific fact about the way people think about corporations and environmental damage' (192). He must have worried about whether recent corporate scandals or social concern for the environment might have skewed his results. The second experiment was structurally identical to the first, but includes vignettes about sending soldiers to their possible doom. We fail to see why this would be any less socially loaded. If Knobe wanted to protect against current social concerns skewing his results, he would need at least one vignette that was not socially loaded. We describe the second experiment below.

In Knobe's second experiment he surveyed 42 people spending time in a Manhattan public park. They were again assigned randomly into 'harm' and 'help' conditions and given the following vignette:

> A lieutenant was talking with a sergeant. The lieutenant gave the order: 'Send your squad to the top of Thompson Hill.'
>
> The sergeant said: 'But 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.'
>
> The lieutenant answered: '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.'
>
> 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.

In the help condition, the difference in vignette is significant. So we reproduce it below.

> A lieutenant was talking with a sergeant. The lieutenant gave the order: 'Send your squad to the top of Thompson Hill.'
>
> The sergeant said: '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!'
>
> The lieutenant answered: '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.'
>
> 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.

Again subjects were asked to determine blame (in the harm condition) or praise (in the help condition), on a scale from 0–6 and to say whether the lieutenant intentionally placed the soldiers in the line of fire (harm condition) or moved them out of the line of fire (help condition). And again the results were similar. In the harm condition, 77% said that the actor intentionally placed the soldiers in the line of fire. In the help condition, 70% said the actor did not intentionally move them out of the line of fire. The results where highly statistically significant. So again, why the asymmetry?

3. *Knobe's explanation*

Knobe reported that overall subjects said 'the agent deserved a lot of blame (with a mean of 4.8 on the 0–6 scale) in the harm condition, but very little praise (mean of 1.4) in the help condition, and the total amount of praise or blame … was correlated with their judgments about whether or not the side effect was brought about intentionally' (7). In other words, Knobe surmised that the asymmetry in praise and blame correlated well with the asymmetry in judgments of the intentionality of actions. '… they seem considerably more willing to say that a side effect was brought about intentionally when they regard the side effect as bad than when they regard it as good.'

That is where Knobe left the matter in this particular paper.[2]

Our explanation: A defence of view I

Interestingly, Knobe does not conclude from his data that the folk concept of intention and intentional action does not conform to View I above.[3] It is easy to see how one could. For example, in the harm condition it appears that the folk are judging both that the chairman does not intend to harm the environment *and* that he does intentionally harm the environment—in clear violation of View I (that intentionally doing A requires the intention to do A). Were one to conclude the folk concept of intentional action does not conform to View I from this data, the conclusion would be premature. We will explain why.

There are at least two ways to interpret Knobe's data. One way is that his surveys are accessing a clearly articulated core folk concept of intentional action. Another way of interpreting the data is that his surveys are accessing not an articulated core folk concept of intentional action, but a clear folk concept of the pragmatic features of intentional language. By

[2] He does say more in a later paper (Knobe 2003b), and we say more about that in another longer paper. In the later paper Knobe argues for a folk core concept that accords best with View II.

[3] Knobe (2003b) does draw conclusions that imply this later.

'pragmatic' we mean to include judgments the folk may make due to social context that may not be part of the semantic content of a sentence or judgement. For example, Mele (2001) gives the example that if Tom says 'I don't desire to see Bill today' the folk may judge that Tom desires not to see Bill today. Of course, that does not follow, but may be inferred for well-known Gricean reasons. As Paul Grice (1989) pointed out, conversation can lead to such implications when one is not being fully informative. If Tom did not want his audience to believe that he wanted to avoid Bill, he should have said more. In normal conversation, this sentence is a way of implying that one wants to avoid Bill, other things being equal. Similarly, when one exclaims 'you did that on purpose' or 'you did that intentionally,' one may be conversationally implying blame, but blame is not part of the semantic content (or core concept) of doing something on purpose (intentionally). Furthermore, we would point out that Gricean implicatures are cancellable. When one implies blame by saying 'you did that intentionally' one may cancel the implicature by adding 'but, of course, it is okay to do that action intentionally'. So this feature of intentional language is pragmatic and not part of the semantic core of the concept of intention or intentional action.

We are inclined to think that the folk do not have a clearly articulated core concept of intentional action. By that we mean they do not have anything approaching a theory of the mental mechanisms that make an action intentional (nor the counterfactual conditions necessary to link intention to intentional action). Normally there would be no need for such an articulated concept. Indeed, philosophers and cognitive scientists are just now articulating such mental mechanisms. While a fully articulated core concept of intention and intentional action is not necessary in daily life, a full grasp of the pragmatics of intentional language is.

In support of our claim that folk lack a fully articulated core concept of intention and intentional action, we point out that Malle & Knobe (1997) did an elaborate survey of the folk concept of intentional action. They found there to be at least five aspects to intentional action in the minds of the folk: belief, desire, skill, intention, and awareness. They also found that no subjects indicated all five aspects (and the missing item kept changing). This supports our view that the folk do not normally possess a clearly articulated theory of the mental mechanisms of intentional action.[4]

However, the folk do possess a very clear notion of the pragmatic features of intentional action and talk of intentional action due to the role of

[4] If the folk always got the same 4 out of 5 features, then we might be willing to accept that there was something approaching a universal folk concept. The fact that there was significant variation in the missing features suggests there may be no single universal folk concept of intentional action.

talk of intention in social praise and blame. Good actions deemed intentional are more highly praised (and encouraged). Bad actions deemed intentional are more severely blamed (and discouraged). The praise and blame associated with intentional action is part of the pragmatics of the concept, not part of the core. This is because the truth conditions for 'S did act A intentionally' do not include praise or blame. It is not necessary for act A to be good or bad for the action to be intentional.[5] However, the folk may associate intentionality with judgments of praise and blame owing to social or evolutionary pressure (Cosmides & Tooby 1994).[6] Folk may be more inclined to judge 'intentional' an act they want to strongly blame and discourage. We believe that something like this is a very plausible explanation of Knobe's findings.

We suggest that if presented with options consistent with View I above, the folk would be as likely to select those options.[7] In particular, in Knobe's vignette, if the folk had been given two options:

(a) The chairman harmed the environment intentionally
(b) The chairman knowingly incurred the risk of harming the environment

we believe that people would be at least as likely to choose statement (b) as (a). Why is this relevant? It is because with choice (b), the folk can still strongly blame and discourage acts of the type committed by the chairman. It allows them to express their disapproval in a way consistent with the chairman's not intending to harm the environment.[8] If the chairman literally 'does not care at all about the environment' then he does not intend to help or harm it.

We suspect that what is going on in the minds of the folk is that they disapprove of the chairman's indifference to the harm of the environment. They want to blame that indifference and they know that their blame is stronger and more effective at discouraging such acts, if the chairman is said to have done the action *intentionally*. They associate blame with intentional action (and 'blame' with 'intentional'). They likely do not consider whether the chairman actually intends to harm the environment or not. If it were pointed out to them that they were judging that one could do an action intentionally without intending it, it may well confront them with a cognitive disconnect and inconsistency.[9] If they chose option (b) above,

[5] A morally neutral act such as setting one's watch can be perfectly intentional and yet be worthy of neither praise nor blame.

[6] Consider their 'cheater-detection' modules which work on purposive behaviour.

[7] We plan to test this at a later date and present the results in a longer paper.

[8] Mele (2001: 40) makes a similar sort of suggestion, which Knobe (2003b) pursues.

[9] We plan to conduct surveys in which we test to see whether subjects display cognitive inconsistency when confronted with it. The results will be discussed at length in a further paper on the role of surveys in accessing folk notions.

there would be no such cognitive disconnect or inconsistency. Hence, we think that subjects who choose (a) are likely not accessing an articulated core concept of intention or intentional action at all. They are more likely accessing the pragmatic features of the intentional talk. If they did access a core concept and considered consistency, we think they would be inclined to choose (b).

One of the stunning features of Knobe's study is the asymmetry of judgments. While the folk may judge that the chairman (and lieutenant) intentionally acted in the harm condition, they judged that they did not intentionally act in the help conditions. Yet these conditions are structurally isomorphic. What could explain this asymmetry?

Since subjects judge the actions to be done intentionally in the harm conditions, why not in the help conditions? For pragmatic reasons, in the help conditions, the folk may find the attitudes of the chairman and lieutenant so despicable that to say their actions were 'intentional' would be to praise them. The language of the harm conditions seems natural (if uncaring), but the language in the help condition seems highly strained. We cannot picture a lieutenant saying such things as 'look I know we will 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.' We understand that Knobe wanted to keep the vignettes in the help and harm conditions parallel. However, in the help conditions when one says something to the effect that either 'I don't care if I help the environment' or 'I don't care if I save the soldier's lives' there is something pragmatically odd about these utterances. Subjects might wonder why the actors don't care if the good consequence came along with what the actors did intend. Subjects surveyed might even take this indifference in the help condition to express a negative attitude about the good side effects. So we worry that this pragmatic oddity of the vignettes might even be more likely to skew Knobe's results than their socially loaded aspects.

Not wanting to praise those who are indifferent to good outcomes, the folk are understandably reluctant to deem the agent's acts to be intentional. Pragmatics may thus be able to help explain why the actions were not judged to be intentional (without importing praise or blame into the folk core concept of intentional action). Pragmatic forces are at work in the help condition, and they yield results consistent with View I. So in the help condition, people may be making the right judgment, but for the wrong (pragmatic) reason. Or, owing to pragmatics in play in the help condition, the folk may well see that if one is indifferent to the outcome of an action, one is *not intending* that outcome, and not doing the action intentionally (consistent with View I). If this is the case, the real mystery in Knobe's results is why subjects judged the actions to be intentional in the harm conditions. To explain this, we have argued that it is due to the pragmatics of intentional language and blame. Judging the actions to be 'intentional' in the

harm condition pragmatically implies strengthened blame. The subjects surveyed want to levy blame and they are likely not doing a mental check for consistency upon an articulated core concept of intention or intentional action. That is, they are likely not accessing an articulated core concept of intention or intentional action at all. So we suspect that the pragmatic forces at work in the harm condition do run against View I, but mainly because no clearly articulated core concept of intention or intentional action is being consulted.

Alternatively, Knobe did not ask the folk whether the actors *intended* to harm or help the environment (rescue or put in harm's way the soldiers). It *is at least possible* that in the minds of the folk, the actors *did intend* the respective outcomes. In that case, in agreement with Malle & Knobe (1997) we might explain the asymmetry as due to the fact that

> … people may distinguish between intentions and doing something intentionally more for positive behaviours, than for negative behaviours, because it is easy (and common) to have positive intentions but harder to fulfill them intentionally, whereas a person's negative intention is already deviant (and threatening to others) even before fulfilling it intentionally. (116)

The subjects may blame in the harm condition on the basis of the negative attitude of the actors alone. The subjects may take the indifference to be an intention to harm. Whereas, they do not take the indifference in the help condition to be an intention to help. Since this explanation is also consistent with View I, it offers another way of explaining the asymmetry of Knobe's data without abandoning View I.

5. *Conclusion*

In conclusion, we have argued that Knobe's surveys do not discredit View I. We have offered an alternate explanation of his data and pointed out some possible problems with the language of his vignettes. We believe that it is more likely that his surveys are accessing pragmatic features of intention and intentional language than an articulated core concept of intention and intentional action. We do think his experiments are important and the asymmetry is worthy of further investigation, but we think it would be premature to draw any conclusion about the truth of Views I or II at this time, based upon his experiments.[10]

[10] We thank Joshua Knobe for helpful conversation, comments, and for bringing his important results to our attention. We thank Al Mele and Michael Stingl for useful conversation and helpful comments, as well. Finally, special thanks to the University of Delaware's Office of Undergraduate Research for support of this joint research project.

University of Delaware
Newark, DE 19716, USA
fa@udel.edu

References

Adams, F. 1986. Intention and intentional action: the simple view. *Mind and Language* 1: 281–301.
Adams, F. 1997. Cognitive trying. In *Contemporary Action Theory*, vol. 1, ed. G. Holmstron-Hintikka and R. Tuomela, 287–314. Dordrecht: Kluwer.
Bratman, M. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
Cosmides, L. and J. Tooby. 1994. *The Adapted Mind*. Oxford: Oxford University Press.
Grice, H. P. 1989. *Studies in the Ways of Words*. Cambridge, MA: Harvard University Press.
Harman, G. 1976. Practical reasoning. *Review of Metaphysics* 29: 431–63.
Knobe, J. 2003a. Intentional action and side effects in ordinary language. *Analysis* 63: 190–94.
Knobe, J. 2003b. Intentional action in folk psychology: an experimental investigation. *Philosophical Psychology* 16: 309–24.
Malle, B. and J. Knobe. 1997. The folk concept of intentionality. *Journal of Experimental Social Psychology* 33: 101–21.
Malle, B., L. Moses and D. Baldwin, eds. 2001. *Intentions and Intentionality: Foundations of Social Cognition*. Cambridge, MA: MIT/Bradford.
McCann, H. 1986. Rationality and the range of intention. *Midwest Studies in Philosophy* 10: 191–211.
Mele, A. 1992. *Springs of Action*. Oxford: Oxford University Press.
Mele, A. 2001. Acting intentionally: probing folk notions. In *Intentions and Intentionality: Foundations of Social Cognition*, ed. B. Malle, L. Moses and D. Baldwin, 27–43. Cambridge, MA: MIT/Bradford.
Mele, A. and S. Sverdlik. 1996. Intention, intentional action, and moral responsibility. *Philosophical Studies* 82: 265–87.

Intention, intentional action and moral considerations

Joshua Knobe

Adams and Steadman (2004) make a number of important criticisms of my work on the concept of intentional action. It seems to me that some of these criticisms are valid. The evidence I presented earlier is indeed open to alternative explanations, and it would be premature to infer, solely on the basis

Analysis 64.2, April 2004, pp. 181–87.