

were right the present reasoning would be no better than the reasoning which Brueckner criticises. For the present reasoning still relies on (**). I reply that there is nothing absurd about the idea that just about any sound philosophical and mathematical argument fails the negation test. It would be absurd to say that just about any such argument is trivial. But failure in the negation test is not enough to make an argument trivial. An argument fails the negation test if and only if were a normally intelligent person to want the argument to have a false conclusion, reflection would destroy his attitude. And an argument is trivial only if it fails this test. But there is another necessary condition for the triviality of an argument: were a normally intelligent person to want the argument to have a *true* conclusion, reflection would destroy his attitude. Sound demonic arguments do doubtless meet this further necessary condition. But we cannot infer from (**) that all or most other sound philosophical or mathematical arguments are in the same boat. It is perfectly possible to have a reflectively indestructible desire for the truth of what one believes to be necessarily true.

*University of Liverpool,
Liverpool L69 3BX, UK
N.M.L.Nathan@liverpool.ac.uk*

References

- Brueckner, A. 2003. Not wanting to know. *Analysis* 63: 250–56.
Nathan, N. M. L. 2001. *The Price of Doubt*. London: Routledge.

Intentional action and moral considerations: still pragmatic

FRED ADAMS & ANNIE STEADMAN

1. Introduction

Adams & Steadman (2004) argued that the core concept of intentional action does not involve moral considerations, but that moral considerations are involved in the pragmatic implications of intentional talk. Joshua Knobe (Knobe 2003a, 2003b) discovered data showing an asymmetry in folk judgments about intentional action. Persons answering Knobe's questionnaires judge actions to be performed intentionally when

the actions involved harming the environment, but not when exactly similar actions involved helping the environment. We suggest that this is because one's saying that an action is done 'intentionally' or 'on purpose' pragmatically implies moral blame – even if blame is not part of the literal, semantic content of 'S did A intentionally.' Where respondents judge actions to be performed intentionally, it is likely that respondents want to assign blame to the persons harming the environment. We find insufficient evidence in Knobe's data to conclude that folk concepts of intentional action contain a slot to record moral considerations of actions as part of a core concept. We maintain that the truth conditions of 'S did A intentionally' do not involve moral considerations, but that one's saying this sentence or thinking it could have moral considerations as pragmatic features that are not part of the semantics. We concluded that pragmatic (not semantic) features of intentional language likely explain Knobe's data.

2. *Knobe's new data*

Knobe (2004) prepared new surveys that do not include the words 'intention' or 'intentional' or 'on purpose', to test the hypothesis that these words and their pragmatic implications explain the asymmetry of folk judgements. He found the same asymmetry of judgment as before and concluded that the asymmetry cannot be explained by pragmatic features of intention talk. We disagree with his conclusion and will explain why.

To quickly review the harm and help vignettes that Knobe presents to folk in his surveys, the *harm vignette* is 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.

The *help vignette* is 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, and it will also help the environment.'

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

They started the new program. Sure enough, the environment was helped.

In the original surveys, Knobe asked respondents (who were equally divided into the 'harm' and 'help' conditions) to judge whether the chairman acted intentionally. In the harm condition, 82% said the chairman harmed the environment intentionally. In the help condition, 77% said the chairman did not help the environment intentionally. Knobe took the results to show that the folk concept of intentional action involves moral considerations at its core.

In response to our claim that pragmatic features of intentional talk also explain the asymmetry, Knobe revised his procedure. In his first survey Knobe asked respondents 'Did the chairman harm (or help) the environment *intentionally*?' In his new survey of 77 respondents, Knobe asked instead whether it *sounds right to say*, 'The chairman harmed (or helped) the environment in order to increase profits?' He found:

Subjects answered this question by providing ratings on a scale from -3 ('sounds wrong') to +3 ('sounds right'), with the 0 point marked 'in between'. The average rating for subjects in the harm condition was +0.6; the average help condition was -1. This difference is statistically significant, $t(77) = 2.65$, $p = 0.01$. (Knobe 2004: 184)

Notice that the new survey questions do not employ the words 'intention' or 'intentionally'. Knobe admits, 'Adams and Steadman are right ... that if a person says, "The chairman did not harm the environment intentionally," there may be an implicature that the chairman was not to blame for harming the environment' (184). However, he adds

But no such implicature arises when a person says, 'It [does not_{A&S}] sound wrong to me to utter the sentence "The chairman harmed the environment in order to increase profits".' (184)

Therefore, he concludes that it seems unlikely that 'the difference between people's responses to the harm vignette and their responses to the help vignette is entirely due to pragmatic factors' (184).

3. *Our new reply*

Our view is that pragmatic features of intentional talk still explain Knobe's findings. Knobe now believes that he has taken pragmatic considerations out of play, by not using the words 'intention' or 'intentional' in his questions on his surveys. This cannot be right because 'in order to' language is also *fully intentional*!

Consider the sentence 'S did A intentionally.' Clearly this is intentional talk. It presupposes a purpose or goal that constitutes the reason why S

did A. S might have done A in order to do B. Or S might have done A for its own sake, but, being intentional, there is some reason why S did A. So with the understanding that A might be identical to B (when doing an action for *its own sake*), there is a mutual entailment between sentences of the form 'S did A in order to B' and those of the form 'S did A intentionally.'¹

Knobe's new wording still is *intentional talk*, and is ripe for pragmatic features.

Now consider Knobe's 'S did A in order to B.' When he asks respondents whether the chairman harmed the environment in order to increase profits, it is implicit that this is an intentional context. If it is not accidental that S did A, then B is the goal of doing A or the reason for doing A. So our point is that 'S did A in order to B' is *just as clearly intentional language* as is 'S did A intentionally.'

Notice also that Knobe is asking the respondents about 'how it sounds' to say the chairman harmed or helped the environment. This surely suggests that it will be possible for *one's saying* something to have pragmatic features that the semantic content of *what is said* does not (which is our view). Even Knobe's questions falls in line with our preferred explanation.² For there may be pragmatic features of one's saying 'S did A in order to B' that are not part of the semantic content of the thought or utterance.

4. *Our old reply*

In our previous paper, we maintained that Knobe should have given subjects a choice between (a) or (b):

- (a) Did the chairman harm the environment intentionally? (Yes or no)
- (b) Did the chairman allow harm to the environment knowingly (but not intentionally)? (Yes or no)

We believe that it is equally likely that subjects would answer yes to (b) as to (a). We do not think that most subjects have a detailed folk concept

¹ We say this with two restrictions. First, we allow that it is possible that A = B. Second, we will just stipulate that we are talking about cases of non-deviance. The mutual implication may not hold in cases of causal deviance. However, none of Knobe's examples involve deviant chains, and no part of his view involves such cases. So this stipulation will not weaken any point Knobe is defending in his new paper.

² We cannot help but wonder why Knobe did not simply ask subjects 'Did the chairman harm the environment in order to raise profits?' or 'Did the chairman help the environment in order to raise profits?' Why ask about how it sounds, if you are trying to take pragmatics out of play?

of what it takes to do something intentionally. They may however accept that one must intend an action to do it intentionally. They also may accept that one can be held morally responsible for actions they did not intend, but knowingly allowed. Offering subjects a choice of answer (b) provides an opportunity to emphasize that the chairman's action was not intentional, but that he is still morally responsible for his action. Subjects might want to choose (b) because the chairman says that he 'doesn't care at all' about what happens to the environment. If he doesn't care *at all* then he neither intends to harm nor to help the environment. So (b) may be the best choice, given the language of the vignette.

We would make the same point with respect to Knobe's question about whether 'S did A in order to B.' That is, we would give the subjects the option of choosing between (a') and (b'):

- (a') S knowingly allowed A in order to do B (but did not intend A).
- (b') S did A intentionally in order to do B.

Given this choice, subjects would have to confront the option of assigning blame in more than one way. We believe subjects would be at least as likely to accept that the chairman knowingly allowed the environment to be harmed in order to increase profits (but did not intend harm) as to accept that the chairman intentionally harmed the environment (without intending to harm it) in order to increase profits. Given that the chairman says he doesn't care about the environment (neither to harm it nor to help it), (b') would be an odd choice *unless* one chose it for the pragmatic implicature that one is morally more responsible for things said to be done intentionally. We suspect that respondents to Knobe's questionnaire want to levy maximum blame upon the chairman by their responses, and that they choose the response with the strongest pragmatic link between intentionality and moral blame.

So, as in our earlier paper, we believe that it is still pragmatics (not semantics) of intentional talk that explains Knobe's discovered asymmetries of judgement. Subjects choose the intentional language that will levy maximum moral clout, and they do so because of the pragmatics of the intentional talk, not the semantics. Further, talk of 'S doing A in order to B' is highly *intentional*, despite the absence of the words 'intention' or 'intentional'.

5. *Intention and intentional action: it's what's in the head that counts*

Nothing in Knobe's earlier experiments tests for whether subjects thought the chairman had an intention to harm the environment. In Knobe's vignettes, the chairman seems to intend neither to harm nor to help the environment. In the harm condition, many subjects say the chairman harmed the environment intentionally. They may be attributing to the

chairman the intention to harm the environment. Or, consistent with our pragmatic hypothesis, they may just be blaming him for his indifference to what happens to the environment. In the help condition, it is clear to the subjects that the chairman does not intend to help the environment, and they don't want to praise him for his indifference (by saying he helped the environment intentionally).

In a follow-up experiment, Knobe explicitly asked subjects about the chairman's intentions. This time Knobe surveyed 63 subjects randomly assigned to the *help* or *harm* conditions. Within each of these conditions, subjects were randomly divided into the *intention* or *intentionally* conditions. Subjects in the *intention* condition were asked whether or not the chairman intended to harm/help the environment. Subjects in the *intentionally* condition were asked whether the chairman helped/harmed the environment intentionally.

The results were as follows:

	<i>Harm</i>	<i>Help</i>
Intentionally	87%	20%
Intention	29%	0%

As before, most people felt that the chairman acted *intentionally* in the harm condition, but not in the help condition. The difference was statistically significant. Few people said that the chairman had the *intention* to harm the environment. Thus, within the harm condition there was a significant difference between those who thought the chairman had the 'intention' to harm the environment and those who thought that the chairman acted 'intentionally'. Again the difference is statistically significant.

Knobe considers this result to be 'striking' (185) and to raise 'interesting questions about the relation between people's concept of intention and their concept of intentional action' (186). We agree, but not for the reasons that Knobe thinks.

First, in the harm condition, nearly 30% of people do judge the chairman *to have* the intention to harm the environment, even though the chairman professes indifference.³ So these subjects may also have judged

³ See also Shaver (1985, Chapter 5) who notes that behaviour by itself may be sufficient for attribution of responsibility and intentionality of an act. Also, in 'attribution theory' there is a tendency to infer 'dispositions' to intend and intentionally do harm on the part of subjects who knowingly do harm. This too can be part of the explanation of the responses to Knobe's questionnaires and is consistent with our pragmatic hypothesis. Someone known to be indifferent to harming the environment is likely to be disposed to intentionally harm it and likely to be attributed intentional harming of it.

that he intended to harm the environment. At most we would have to employ our pragmatic hypothesis to explain why the other 60% would judge the chairman to harm the environment intentionally without intending to harm it.⁴

Second, *why would* only 30% say the chairman intended to harm the environment, if nearly 90% said he harmed the environment intentionally? Knobe would have us think that it is because the folk concept of intentional action essentially involves a slot or feature for moral considerations. He would have us believe that if an action is morally relevant, that figures in one's judgment of whether or not the action is intentional. If true, why don't subjects judge the chairman to have a morally positive intention to help the environment in the help condition (where his behaviour causes a good outcome to the environment)? The respondents no doubt think it is morally good to help the environment *and* the chairman *did help the environment*. So why don't they attribute intentionality to his helping? We think that a more plausible explanation is that, to attribute intentionality to the chairman's action, the folk may be looking more at what they deem to be the chairman's behaviour. The vignette says that he instituted the new policy that resulted in harm to the environment. So his *behaviour* was the same as that of someone who harmed the environment intentionally, and the subjects disapprove of that outcome. At this point we believe pragmatic features of intentional language come into play in the subjects' judgments.

Why don't they attribute *intentions* to the chairman? In the help condition, no one attributed an intention to help the environment, and in the harm condition only about 30% attributed an intention to harm. If there really is a slot in folk judgment for moral considerations and the respondents judge the chairman's actions to be morally bad, why did not *all of the respondents* judge the chairman to have intended harm to the environment? We suggest that it is because to judge that the chairman intended to harm the environment, is to say something about what is going *through his head*. It is to attribute beliefs, desires, and plans to the chairman. One may not be able to tell, just from the behaviour of another (or its description in a vignette) how that person represents his or her goals or plans. Subjects may have a harder time saying (determining) whether the chairman actually planned, or desired to harm the environment. This too is why we say that it easier for subjects to use the phrase 'S intentionally did A' to lay blame on the actions of the chairman. From his behaviour (verbal and otherwise) it is clear that he knowingly caused something of which the respondents disapprove. What is less clear is what went through the chairman's head (mind).

⁴ Assuming Knobe's results are representative.

Norman Malcolm (1972) once gave the following view. We *could say* his dog thinks that a cat ran up a tree because the dog was chasing the cat, went to the tree, barked up the tree, and stood looking into its branches. The behaviour was clearly intentional and seemed clearly directed at the cat. Hence it is easy to judge the behaviour intentional. However, Malcolm claimed that *one could not say* that the dog *had the thought that* the cat ran up the tree. Why? What's the difference? The latter requires knowing what goes through the head of the dog. How does it think about the cat, the tree, and running up something? Unless we know these things, we don't know the specific content of the dog's thoughts.

Similarly, it is easier for folk to make judgments about the intentionality of the *behaviour* of the actors in the vignettes than to judge the actual contents of the actors' *minds*. Thus, we are not surprised that there are low numbers (even zero) in the help and *intention* condition, relative to the *intentional* condition. Unlike Knobe, we maintain that his data are insufficient to tell whether folk concepts of intentional action involve semantic slots for moral considerations or whether their judgments are being guided by pragmatic aspects of intentional talk.

6. Conclusion

We continue to be impressed by Knobe's results. They are 'striking,' and they need to be explained. We are also impressed with Knobe's ingenuity in designing experiments attempting to show that folk judgments about intentional actions are sensitive to moral dimensions. He believes his data show that core folk concepts entail that actions can be intentional, though not intended. If true, there is a semantic marker for whether an action is morally right or wrong in one's concept of intentional action. And this marker can override other markers for whether the action is intended. If true, this would indeed be surprising.

However, we continue to believe there is an equally plausible explanation of the data that does *not* support that conclusion. Our view is that the folk do not have a fully articulated concept of intentional action. They do know that beliefs, desires, goals, planning, and intentions are relevant to intentional action. They also know that there is a close connection between what one does on purpose and one's moral praise or blame. Furthermore, they utilize pragmatic features of intentional talk to attribute moral praise or blame to actions. Saying that one did something 'intentionally' or 'in order to' suit a purpose can emphasize the action's moral properties, independently of whether one has good information about what is going through an actor's mind (what he or she actually intends). Thus, we do not believe that moral dimensions of judgments about an

action determine whether or not the action *is indeed* an intentional action. We suggest instead that any appearance to that effect in Knobe's data is equally well explained by appealing to the pragmatic features of intentional language. Indeed, we suspect that if Knobe were to put on his questionnaire 'Is it possible to do A intentionally without intending to do A?' the folk would experience cognitive dissonance. They would likely hesitate to embrace *both* that the chairman intentionally harmed the environment *and* that he did not intend to harm the environment.⁵ Confronting the tension, they may jump at the chance to accept that the chairman allowed the environment to be harmed knowingly, but did not do it intentionally (though he may have intentionally incurred the risk of causing environmental harm). Hence, we still maintain that the connection in the mind of the folk between intentional action and moral concern is likely to be pragmatic (not semantic).⁶

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

References

- Adams, F. and A. Steadman. 2004. Intentional action in ordinary language: core concept or pragmatic understanding? *Analysis* 64: 173–81.
- Knobe, J. 2003a. Intentional action in folk psychology: an experimental investigation. *Philosophical Psychology* 16: 309–24.
- Knobe, J. 2003b. Intentional action and side-effects in ordinary language. *Analysis* 63: 190–94.
- Knobe, J. 2004. Intention, intentional action and moral considerations. *Analysis* 64: 181–87.
- Malcolm, N. 1972. Thoughtless brutes. *Proceedings and Addresses of the American Philosophical Association* 5–20.
- Shaver, K. 1985. *The Attribution of Blame: Causality, Responsibility, and Blameworthiness*. New York: Springer-Verlag.

⁵ Notice that Knobe has subjects divided either into the 'intentionally' condition or the 'intention' condition, but not both. So none have had to face the cognitive dissonance we are pointing out.

⁶ We'd like to thank Al Mele for useful comments on an earlier draft. We also thank the University of Delaware Office of Undergraduate Research for support.