Intentional action and side effects in ordinary language

JOSHUA KNOBE

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.

Here it appears that, although the chairman foresaw that x would result from his behaviour, he did not care either way whether x actually occurred. Let us say, then, that x was a 'side effect' of his behaviour. The question I want to address here is: Shall we say that the chairman brought about this side effect *intentionally*?

This question goes to the heart of a major controversy regarding the proper analysis of the concept of intentional action. So, for example, on Alfred Mele's (2001) analysis, it is always wrong to say that a side effect was brought about intentionally.[1] By contrast, on Michael Bratman's (1984; 1987) analysis, there are circumstances under which side effects can truly be said to have been brought about intentionally. Numerous other authors have come down on one side or the other of this issue.

Now, when we encounter a controversy like this one, it can sometimes be helpful to ask ourselves what people would ordinarily say about the situation under discussion. Would people ordinarily say that the side effects of a behaviour were brought about intentionally? Clearly, ordinary language does not here constitute a court of final appeal. (Even if it turns out that people ordinarily call side effects 'intentional', we might conclude that they are truly unintentional.) Still, it does seem plausible that the examination of ordinary language might provide us with some useful guidance about difficult cases like this one.

In an earlier publication, the experimental psychologist Bertram Malle and I provided empirical support for the conclusion that people only con-

[1] Mele (2003) now retracts this view in response to an earlier version of the present paper.

ANALYSIS 63.3, July 2003, pp. 190–94.

sider an effect to have been brought about 'intentionally' when the agent was specifically trying to bring about that effect (Malle & Knobe 1997). I now think that this conclusion was too hasty. The truth is that a person's intuitions as to whether or not a given side effect was produced intentionally can be influenced by that person's attitude toward the specific side effect in question (Harman 1976). Thus, it would be a mistake to ask for a general answer to the question as to whether people will think that an agent intentionally brought about 'some side effect *x*'. People's judgements depend in a crucial way on what *x* happens to be. In particular, it makes a great deal of difference whether they think that *x* is something good or something bad.

1. *First experiment*

To test this hypothesis, I conducted a simple experiment. Subjects were 78 people spending time in a Manhattan public park. Each subject was randomly assigned to either the 'harm condition' or the 'help condition'. Subjects in the harm condition read the following vignette:

> 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.

These subjects were then asked to determine how much blame the chairman deserved for what he did (on a scale from 0 to 6) and to say whether they thought the chairman *intentionally* harmed the environment.

Subjects in the help condition received a vignette that was almost exactly the same, except that the word 'harm' was replaced by 'help':

> 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.

These subjects were then asked to determine how much praise the chairman deserved (on a scale from 0 to 6) and whether they thought the chairman *intentionally* helped the environment.

As predicted, the two conditions elicited two radically different patterns of responses. In the harm condition, most subjects (82%) said that the agent brought about the side effect intentionally, whereas in the help condition, most subjects (77%) said that the agent did not bring about the side effect intentionally. This difference was highly statistically significant, $\chi^2(1, N = 78) = 27.2, p < .001$.

2. *Second experiment*

Of course, we should be reluctant to reach any general conclusions on the basis of just one experiment. 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. To test the generality of the effect, I therefore ran a second experiment. The vignettes used in this second experiment had the same basic structure as those used in experiment 1, but the story was transposed from a corporate setting (with a chairman affecting the environment) to a military setting (with a lieutenant affecting a squad of soldiers).

Subjects were 42 people spending time in a Manhattan public park. Once again, subjects were randomly assigned to either a 'harm condition' or a 'help condition.' Subjects in the harm condition were asked to read 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.

These subjects were then asked to determine how much blame the lieutenant deserved for what he did and whether or not he intentionally put the soldiers in the line of fire.

Subjects in the help condition received a quite similar 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: '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.

These subjects were then asked to determine how much praise the lieutenant deserved for what he did and whether or not he intentionally took the soldiers out of the line of fire.

Once again, the two conditions elicited two radically different patterns of responses. In the harm condition, most (77%) subjects said that the agent brought about the side effect intentionally, whereas in the help condition most (70%) subjects said that the agent did not bring about the side effect intentionally. This difference was statistically significant, $\chi^2(1, N = 42) = 9.5, p < .01$.

3. *Explaining the results*

Why do people respond so differently to vignettes that seem, at least in certain respects, to be so similar? Here subjects' ratings of praise and blame may provide an important clue. I therefore combined the praise and blame ratings from the two experiments and ran a new series of tests.

Overall, subjects said that the agent deserved a lot of blame in the harm condition ($M = 4.8$) but very little praise in the help condition ($M = 1.4$), $t(120) = 8.4, p < .001$, and the total amount of praise or blame that subjects offered was correlated with their judgements about whether or not the side effect was brought about intentionally, $r(120) = .53, p < .001$.

In other words, there seems to be an asymmetry whereby people are considerably more willing to blame the agent for bad side effects than to praise the agent for good side effects. And this asymmetry in people's assignment of praise and blame may be at the root of the corresponding asymmetry in people's application of the concept *intentional*: namely, that they seem considerably more willing to say that a side effect was brought about intentionally when they regard that side effect as bad than when they regard it as good.[2]

Princeton University
Princeton, New Jersey 08544-1006, USA
jknobe@princeton.edu

[2] I am grateful for comments from Alfred Mele, Alan Leslie and Adam Elga.

References

Bratman, M. 1984. Two faces of intention. *Philosophical Review* 93: 375–405.
Bratman, M. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
Harman, G. 1976. Practical reasoning. *Review of Metaphysics* 29: 431–63.
Malle, B. F. and J. Knobe. 1997. The folk concept of intentionality. *Journal of Experimental Social Psychology* 33: 101–21.
Mele, A. 2001. Acting intentionally: probing folk notions. In *Intentions and Intentionality: Foundations of Social Cognition*, ed. B. F. Malle, L. J. Moses & D. Baldwin, 27–43. Cambridge, MA: M.I.T. Press.
Mele, A. 2003. Intentional action: controversies, data, and core hypotheses. *Philosophical Psychology* (in press).

A probabilistic theory of coherence

BRANDEN FITELSON

1. *The coherence measure* C

Let **E** be a set of n propositions $E_1, \ldots, E_n$. We seek a probabilistic measure $C(\mathbf{E})$ of the 'degree of coherence' of **E**. Intuitively, we want C to be a quantitative, probabilistic generalization of the (deductive) *logical coherence* of **E**. So, in particular, we require C to satisfy the following intuitive desideratum.

(1) $C(\mathbf{E})$ *is*

Maximal (positive, constant)	if the E_i are logically equivalent (and **E** is satisfiable)
> 0	if **E** is positively dependent[1]
0	if **E** is independent[1]
< 0	if **E** is negatively dependent[1]
Minimal (negative, constant)	if all subsets of **E** are unsatisfiable

Desideratum (1) captures the qualitative features that a probabilistic generalization of logical coherence should satisfy – it requires C to respect the extreme deductive cases, and to be properly sensitive to probabilistic dependence (a general notion of probabilistic dependence will be defined precisely, and in a slightly non-standard way, below).

I propose a probabilistic measure of coherence C based on a slight modification of Kemeny and Oppenheim's (1952) measure of factual support F.

[1] See below for definition.

ANALYSIS 63.3, July 2003, pp. 194–99.