

On Praise, Side Effects, and Folk Ascriptions of Intentionality

Thomas Nadelhoffer

Florida State University

Introduction:

In everyday discourse, we often draw a distinction between actions that are performed intentionally (e.g. shutting a car door) and those that are performed unintentionally (e.g. shutting a car door on your finger). This distinction has interested philosophers working in a number of different areas (e.g. action theory, free will, moral responsibility, and the philosophy of law). And while most philosophers agree that the concept of intentional action plays an important role in our folk psychology, there is still wide-scale disagreement about the precise nature of this role. Until recently, because there has been a dearth of empirical data about folk ascriptions of intentional action, the conceptual analyses of intentionality developed by philosophers have been mostly *speculative*. Lately, however, a number of both philosophers and psychologists have begun making a concerted effort to fill in this empirical lacuna.

Joshua Knobe, for instance, has recently published data that suggest that moral considerations affect people's judgments about whether a foreseen yet undesired side effect of an action was brought about intentionally (2003a). More specifically, he claims that people are "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" (2003a: 193). Interestingly, this result does not settle with a study of mine involving folk ascriptions of intentionality in standard cases of action that do not involve side effects (Nadelhoffer forthcoming). The data from my non-side effect experiments suggest that

both positive and negative moral considerations affect people's judgments about intentionality.

There are at least two ways to explain the asymmetry of the results from our experiments: First, it could be that people's judgments about the intentionality of non-side effect actions are sensitive to positive moral considerations in a way that their judgments about the intentionality of side effect actions are not. In this case we would need to explain why negative but not positive moral considerations affect people's judgments concerning the intentionality of side effects, while both types of consideration affect their judgments about non-side effect actions. Second, we could look for another way of explaining the asymmetry (e.g. Adams & Steadman 2004a; 2004b). In this paper I take the latter route by suggesting that Knobe gets disparate results because the two help vignettes of his side effect experiments are not proper praiseworthy analogues to his morally blameworthy harm vignettes. After briefly setting the stage (§I) and discussing some of the data from experiments conducted by both Knobe (2003a) and myself (forthcoming) (§II), I offer an alternative account of his data that explains the asymmetry of his subjects' judgments in the side effect cases (§III). Then, I support my explanation of Knobe's research on side effects with data from a new experiment of my own before concluding that people's judgments about actions – and the side effects of those actions – are affected by both negative and positive moral considerations (§V).

I: Setting the Stage

Humans are quite adept at explaining and predicting the behaviour of one another. These explanations and predictions usually involve taking other people's desires, beliefs, intentions, emotions, reasons, and plans into consideration. As Paul Churchland

suggests, “a rich network of common-sense laws can indeed be reconstructed from this quotidian commerce of explanation and anticipation; its principles are familiar homilies; and their sundry functions are transparent” (1981: 68). On this view, our ability to coordinate and sustain complicated social activities grows out of our having grasped the various relationships that hold between “external circumstances, internal states, and overt behaviour” (Churchland 1981: 68). For present purposes, we shall follow Churchland in calling all of the processes, considerations, and concepts that enable us to grasp these relations our “folk psychology.” In this paper, however, rather than examining the whole web of interrelated concepts that fall under the rubric of “folk psychology,” we will focus primarily on the concept intentional action. But before we examine this concept in detail, we should first briefly discuss the distinction between an agent’s *intentions* and the actions that an agent performs *intentionally*.

Perhaps the easiest way to distinguish intentions from intentional actions is to suggest that the former, unlike the latter, are “in the head,” so to speak. According to one standard account of intentions, an intention has both a motivational and a cognitive component. So, for instance, Wayne Davis claims that, “S intends that p iff S believes that p because he desires that p and believes his desire will motivate him to act in such a way that p” (1984: 54). Moreover, he suggests that a) to believe that p “is equivalent to being more certain that p than non-p” and b) to desire p is “equivalent to preferring p to non-p” (1984: 44-45). On this belief-desire model, intentions are complex psychological states of the agent. But given that intentions are mental states, the question of how they produce external intentional actions immediately arises.

According to the analysis of intentional action that Michael Bratman has dubbed the “Simple View,” in order for an agent *S* to intentionally *p*, *S* must first *intend* to *p* (1984: 377). On this view, in order for my *p-ing* to count as an intentional action, at the time that I *p* my mental states must be such that *p-ing* is among the things that I intend to do (e.g. Adams 1986; McCann 1986; 1991). Despite the intuitive plausibility of this view, not all philosophers have attempted to explain intentional actions in terms of prior intentions. Some have tried to explain them in terms of an agent’s having a *reason* to *p* (e.g. Audi 1985), whereas others have claimed that the most important factor in determining whether an agent performed an action intentionally is whether the agent was *trying* to perform the action (e.g. O’Shaughnessy 1973). But since we are not presently concerned with the causal role that certain mental states play in the production of intentional action, we can leave aside the question of whether intentions, reasons, or “tryings” deserve explanatory pride of place. Instead, we will examine the various conditions within which people ordinarily ascribe intentionality to one another – i.e. with folk ascriptions of intentional action.

Even though philosophers disagree about how intentional actions are produced, everyone seemingly agrees that the distinction between intentional and unintentional action plays an important role in our collective folk psychology. In a number of ordinary situations, the question of whether or not an action was performed intentionally can make a big difference in how we respond to it. So, for instance, if someone accidentally hurts my feelings, I will readily accept her apology insofar as her hurting my feelings was *unintentional*. If, on the other hand, someone purposely, knowingly, and intentionally hurts my feelings, I will be much less likely to excuse her behavior. My different

responses to these two kinds of situations suggest that at least one role that the concept of intentional action plays in our folk psychology is that of fixing blame and praise. And while we can occasionally be held morally and legally responsible for unintentional actions – e.g. cases of negligence or recklessness – intentional actions are more commonly the target of our moral judgments. But the exact nature of this role is not entirely clear.

Consequently, there is a long-standing debate in the philosophy of action concerning the nature and proper role of ascriptions of intentionality. One of the central issues of this debate is whether moral considerations do – or *should* – affect our application of the concept of intentional action. While some scholars suggest that this concept is intimately bound up with moral considerations (e.g. Bratman 1987; Duff 1982; 1990; Harman 1976; Knobe 2003a; 2003b; 2004; Nadelhoffer 2004; forthcoming), others claim that moral consideration should not act expansively on our ascriptions of intentional action. On this latter view, while we may correctly appeal to the intentionality of an action in our attempt to determine someone’s moral or legal responsibility, the converse is not the case – i.e. attributions of blame and praise should not affect our ascriptions of intentional action. Mele and Sverdlik (1996) offer the most well developed and forcefully argued defense of this view (see, also, Butler 1978; Katz 1987). By their lights, philosophers who claim that, “ordinary speakers of English are, to some extent, properly guided by their judgments of moral responsibility in determining what an agent did intentionally,” are wrong about the evaluative nature of *proper* ascriptions of intentional action (Mele & Sverdlik 1996: 270).

Another interesting – and closely related – debate found in the philosophical literature concerns the question of whether a foreseen yet undesired side effect may be brought about intentionally. Consider, for instance, the following example: In the landmark *Smith* case of 1961, jurors in England had to determine the guilt of a man named Smith who had driven a car containing stolen goods in a zigzag course in order to shake off a policeman who had been clinging to the side of the car. When the policeman was finally shaken off, he rolled into oncoming traffic and sustained fatal injuries (*D.P.P v. Smith* [1961] A.C. 290).¹ Since Smith obviously *intended* to both shake off the policeman and to escape, it is clear that his shaking off the policeman and his escaping were *intentional* actions. But what should we say about the unfortunate death of the policeman – an undesired side effect of Smith’s actions? Imagine that you are on that jury and your task is to decide whether Smith intentionally brought about the policeman’s death. At least two questions come to mind: First, can you properly say that Smith intentionally brought about the death of the policeman even though his doing so was a *side effect* of Smith’s escaping that he neither tried nor hoped to bring about? Second, should the fact that this side effect is morally bad act affect your judgment concerning whether it was brought about intentionally? On the surface, it seems that the badness of the consequences of Smith’s actions should be completely *irrelevant* to the question of whether he performed them *intentionally*, but there is growing evidence that these types of moral considerations do affect our ascriptions of intentionality (Knobe 2003a; 2003b; 2004; forthcoming; Nadelhoffer 2004; forthcoming).

¹ For detailed discussions of this case, see, e.g. Finnis (1991); Gorr (1996); Hart (1968); Kenny (1968); Lyons (1976), Oberdeik (1972).

The view that moral features both do and should influence our intuitions concerning the intentionality of side effects has received support in the philosophical literature. Michael Bratman, for instance, claims that a runner may intentionally wear down the soles of his heirloom shoes even though he doesn't intend to do so (1987: 123).

As he says:

I conjectured that our inclination to extend what I do intentionally, in the light of my belief about my sneakers, is grounded in our interest in the ascription of responsibility. Our scheme for classifying actions as intentional is shaped in part by an interest in locating paradigm actions for which agents are to be held responsible. (1987: 136)

On this view, our concept of intentional action is intimately bound up with our notion of responsibility such that whether we say some action p is intentional may partly depend on whether p is something that we either could or should be held responsible for performing. Thus, even though Bratman's runner does not intend to wear down the sole's of his heirloom shoes, to the extent that he knowingly wears them down and can be properly held responsible for doing so, he does so intentionally. In this respect, Bratman is denying the aforementioned Simple View – i.e. the view that in order for S to p intentionally, S must first intend to p .

Gilbert Harman is another philosopher who rejects the Simple View. As he says, “it is a mistake to suppose that whenever someone does something intentionally, he intends to do it” (1976: 433). To support this claim, Harman points to cases involving foreseen yet unintended and undesired side effects. By his lights, for instance, a sniper may intentionally alert his enemies when he shoots his target, even though he neither desires nor intends to alert them (1976: 433). On this view, we can properly say that the sniper intentionally alerted the enemies even though he did not intend to because “in

firing his gun, the sniper knowingly alerts the enemy to his presence. He does this intentionally, thinking that the gain is worth the possible cost. But he certainly does not intend to alert the enemy to his presence” (1976: 433). Harman does not want to say that the soldier intended to alert the enemies because he did not desire to alert them; indeed, the soldier had both a desire and a reason *not* to do so.

Next, Harman contrasts the case of the sniper who alerts his enemies with a case involving a soldier who successfully shoots a bull’s-eye. In this latter example, Harman suggests that if the soldier skillfully makes the shot then his doing so is intentional; whereas if the soldier luckily makes the shot, then his doing so was not intentional. In explaining our conflicting intuitions in these cases, Harman makes the following claim:

The reason why we say that the sniper intentionally kills the soldier but do not say that he intentionally shoots a bull’s-eye is that we think that there is something wrong with killing and nothing wrong with shooting a bull’s-eye. If the sniper is part of a group of snipers engaged in a sniping contest, they will look at things differently. From their point of view, the sniper simply makes a lucky shot when he kills the soldier and cannot be said to kill him intentionally. The same sort of consideration leads us to say that, in firing his gun, the sniper intentionally alerts the enemy to his presence. We say this because the sniper acts in the face of a reason not to alert the enemy to his presence. (1976: 433 – 34)

Thus, on Harman’s view, moral considerations – in this case having reasons not to do something – do and should act expansively on our ascriptions of intentional action.²

Mele and Sverdlik, on the other hand, deny this claim and offer an error theory that explains why the folk may be inclined to *improperly* allow moral considerations to affect their intuitions about whether the foreseen yet unintended and undesired side

² Insofar as both Bratman and Harman believe that moral considerations affect folk judgments concerning the intentionality of the side effects of an agent’s actions, they would likely agree with R.A. Duff’s suggestion that “Ascriptions of...intentional agency belong with ascriptions of responsibility and demands for justification. To say that A brings y about intentionally is to say that he is responsible for its occurrence and may have to justify his action under the description “bringing y about”...the criteria of intentional agency as to a given effect are also the criteria of responsibility for that effect” (1982: 4).

effects of an action are intentionally brought about (1996). By their lights, while the folk may *correctly* assume that Bratman's runner does not unintentionally (i.e. accidentally or unknowingly) wear down his shoes – or that Harman's sniper does not unintentionally alert the enemies to his presence – they incorrectly assume that because the runner and sniper do *not unintentionally* bring about these side effects, it follows that they must have *intentionally* brought them about. Mele and Sverdlik claim that this assumption is false.

On their view, there is a middle ground between unintentionally A-ing and intentionally A-ing, *viz. non-intentionally A-ing* (1996: 274):

Insofar as an agent who is A-ing is neither aiming at A-ing nor trying to A, either as an end or as a means to an end, she is not intentionally A-ing; insofar as an agent is A-ing knowingly and non-accidentally, she is not unintentionally A-ing; and actions that are neither intentional nor unintentional are nonintentional. (1996: 274)

So, for instance, imagine that you habitually whistle in your car every day while driving to work. Moreover, imagine that you are so accustomed to whistling that you neither *intend* to whistle nor *try* to whistle – you simply whistle without giving it any thought or effort at all. Indeed, often you do not even notice that you are doing so. Nevertheless, it would be strange to say that you were *accidentally* whistling. Hence, according to Mele and Sverdlik, you are neither *intentionally* whistling – to the extent that you neither intend nor try to do so – nor *unintentionally* whistling – to the extent that your doing so is not accidental. On their view, you are *non-intentionally* whistling. Having carved out this middle ground, Mele and Sverdlik apply the notion of non-intentionality to Bratman's runner and Harman's sniper. According to this line of reasoning, to the extent that a) neither the runner nor the sniper intended or tried to bring about the side effects of their respective actions, and b) they did not bring about these side effects accidentally, we

should say that they brought them about non-intentionally – i.e. neither intentionally nor unintentionally.³

Leaving aside for the moment the normative question of whether we *should* say that side effects are brought about intentionally, I want to consider whether we actually *do* say so. After all, as Mele has correctly pointed out, any adequate philosophical analysis of intentional action should be “anchored by common-sense judgments” about particular cases (Mele 2001: 27). On this view, one way of testing an analysis of intentional action would be to find out whether it agrees with our pretheoretical judgments and intuitions. And the only method of determining what the majority of non-specialists say about particular cases is to *actually ask them*. Having done so, if we find that an analysis of intentional action entirely fails to settle with folk intuitions, we will be in a good position to suggest that it “runs the risk of having nothing more than a philosophical fiction as its subject matter” (Mele 2001:27). Minimally, any philosopher who offers an account of intentional action that is not anchored by folk judgments would need to offer an error theory that explains how and why the folk are misapplying the concept. But we are getting ahead of ourselves. Before we can talk about how best to interpret and explain data about folk intuitions, we should first look at some recent attempts to get at these intuitions.

II: Empirical Research

A: Folk Ascriptions of Intentional Action in Non-Side Effect Cases

As we saw earlier, there is wide-scale disagreement among philosophers concerning whether moral considerations act expansively on our ascriptions of intentional action. So, in order to answer this question in an empirically verifiable – rather than

³ For a similar discussion of non-intentional action – i.e. action that is neither intentional nor unintentional – see Duff (1990: 77-9) and Mele & Moser (1994).

merely speculative – manner, I conducted some simple psychological experiments and presented the data in an earlier paper of mine (forthcoming).⁴ Subjects were 40 undergraduates, each of whom received the following vignette:

Case 1 (C1):

A nuclear reactor is in danger of exploding. Fred knows that its exploding can only be prevented by shutting it down, and that it can be shut down only by punching a certain ten-digit code into a certain computer. Fred is alone in the control room. Although he knows which computer to use, he has no idea what the code is. Fred needs to think fast. He decides that it would be better to type in ten digits than to do nothing. Vividly aware that the odds against typing in the correct code are astronomical, Fred decides to give it a try. He punches in the first ten digits that come into his head, in that order, believing of his doing so that it ‘might thereby’ shut down the reactor and prevent the explosion. Amazingly, he punches in the correct code, thereby preventing a nuclear explosion and saving thousands of people.

Subjects were then asked the following questions: First, did Fred intentionally punch in the correct numbers? Second, how much praise does Fred deserve for punching in the correct numbers (on a scale from 0 to 6 – 0 being no praise and 6 being a lot)? Third, did Fred intentionally prevent the explosion? And finally, how much praise does Fred deserve for preventing the explosion (on a scale from 0 to 6 – 0 being no praise and 6 being a lot)? The results were as follows:

- Q1: 37% said Fred punched in the correct numbers intentionally.
- Q2: The average praise rating was 3.0 on a 6-point scale.
- Q3: 73% said Fred intentionally prevented the explosion.
- Q4: 4.0 on a 6-point scale.

This shows that even though most of the subjects (*viz.*, 63%) were *not* willing to say that Fred intentionally punched the correct numbers, a surprising majority of them (*viz.*, 73%) *were* willing to say that Fred intentionally prevented the explosion.

⁴ All three of the vignettes in this experiment are borrowed from Mele and Moser (1994) – the main target of my criticisms in the original paper.

Because there is growing evidence that blame has a very pronounced effect on ascriptions of intentional action (see, for example, Knobe 2003a; 2003b; 2004; Nadelhoffer 2004; forthcoming), I wanted to see whether a blameworthy action would similarly affect people's judgments in the Fred example. So, I conducted another survey involving a variation of C1. This time subjects were 40 undergraduates, each of whom received the following case:

Case 2 (C2):

Fred has just been fired from the nuclear power plant. In a desperate fit of anger, he decides to cause the reactor to meltdown. Fred knows that the only way the reactor can be forced to melt down is by punching a certain ten-digit code into a certain computer. Fred is alone in the control room. Although he knows which computer to use, he has no idea what the code is. Fred needs to think fast before the other employees return. Vividly aware that the odds against typing in the correct code are astronomical, Fred decides to give it a try. He punches in the first ten digits that come into his head, in that order, believing of his doing so that it 'might thereby' cause the reactor to meltdown. Amazingly, he punches in the correct code, thereby causing a serious nuclear meltdown and killing thousands of people.

Subjects were then asked the following four questions: First, did Fred intentionally punch in the correct numbers? Second, how much blame does Fred deserve for punching in the correct numbers (on a scale from 0 to 6 – 0 being no blame and 6 being a lot)? Third, did Fred intentionally cause the explosion? And finally, how much blame does Fred deserve for causing the explosion (on a scale from 0 to 6 – 0 being no praise and 6 being a lot)?

The results were as follows:

- Q1: 67% said that Fred intentionally punched in the correct numbers.
- Q2: The average blame rating was 5.23 on a 6-point scale.
- Q3: 83% said that Fred intentionally caused the explosion.
- Q4: The average blame rating was 5.31 on a 6-point scale.

These results show that the *blameworthiness* of Fred in C2 had an even greater effect on subjects' ascriptions of intentional action than did his *praiseworthiness* in C1. After all,

whereas 63% of the subjects in C1 said that Fred *did not* intentionally punch in the correct numbers, 67% of the subjects in C2 said he *did* intentionally punch in the correct numbers. Thus, even though both praise and blame affected the subject's judgments concerning Fred's actions, blame clearly had a more marked effect.

Nevertheless, in an effort to verify that moral considerations – rather than some other variable – fully explain the results of C1 and C2 I conducted another experiment that did not involve a moral component. Subjects were 40 undergraduates, each of whom received the following vignette:

Case 3 (C3):

Imagine that Fred is playing a new kind of lottery machine for \$1,000,000. In order to win, he must type in the correct ten-digit code. Vividly aware that the odds against typing in the correct code are astronomical, Fred pays his \$1 and decides to give it a try. He punches in the first ten digits that come into his head, in that order, believing of his doing so that it 'might thereby' win him the \$1,000,000. Amazingly, he punches in the correct code and wins the lottery!

Subjects were then asked the following two questions: First, did Fred intentionally punch in the correct numbers? Second, did Fred intentionally win the lottery? The results were as follows:

- Q1: 20% said that Fred intentionally punched in the correct numbers.
- Q2: 33% said that Fred intentionally win the lottery.

This data obviously differs drastically from the results of C1 and C2 even though in all three cases, Fred's chances of success is the *exact same* – *viz.*, astronomically small. And to the extent that the only difference between the cases is that Fred's action in C3 is morally neutral rather than blameworthy or praiseworthy, we can reasonably conclude that moral considerations explain the asymmetry of the subjects' responses. If this is correct – and the data suggest that it is – then we have evidence that negative and positive

moral considerations act expansively on folk ascriptions of intentional action – with the former having an even greater effect than the latter.

IIB: Moral Considerations and Side Effect Cases

In another series of similar experiments, Knobe set out to determine whether folk intuitions about the intentionality of foreseeable yet undesired side effects are similarly influenced by moral considerations (Knobe, 2003a). The subjects of his first side effect experiment were presented with a vignette involving either a “harm condition” or a “help condition.” Those subjects who received 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. (2003a: 191)

The subjects were then asked to judge how much blame the chairman deserved for harming the environment (on a scale from 0 to 6) and to say whether they thought the chairman harmed the environment intentionally. 82% of the subjects claimed that the chairman harmed the environment intentionally. Subjects in the help condition, on the other hand, read the same scenario except that the word “harm” was replaced by the word “help.” The subjects were then asked to judge how much praise the chairman deserved for helping the environment (on a scale from 0 to 6) and to say whether they thought the chairman helped the environment intentionally. Only 23% of the subjects claimed that the chairman intentionally helped the environment (2003a: 192).

In another side-effect experiment, Knobe got similar results. This time the subjects received one of the following two vignettes:

Harm Condition:

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 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.

Help condition:

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 taking them 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 moved into the enemy's line of fire, and some of them were saved. (2003a: 192)

Once again, the harm and the help conditions yielded drastically different responses:

77% of the subjects who read the harm condition said the agent intentionally brought about the negative side effect, whereas only 30% of the subjects who read the help condition said the agent brought about the positive side effect intentionally (2003a: 192).

When Knobe combined the praise and blame ratings from the two experiments, he got the following results: Whereas the subjects who were given the harm condition said the agent deserved a lot of blame ($M=4.8$), the subjects who were given the help condition said that the agent deserved virtually no praise ($M=1.4$). Moreover, these results were correlated with their judgments about whether or not the side effect was brought about intentionally (2003a: 193). Thus, Knobe claims that:

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 intentionally: 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. (2003a, p.193)

On the surface, Knobe's conclusion appears to be well supported by his data, but I am now going to suggest that we have good reason to doubt whether this is actually the case.

III: An Alternative Explanation

Because Knobe's harm conditions elicited significantly different results than his help conditions, he concludes that the subjects' ascriptions of intentionality were affected by the moral badness of the side effects of the former but not the moral goodness of the side effects of the latter. Moreover, he claims that whereas bad side effects appear to act expansively on the subjects' ascriptions of blame, good side effects do not have similar effects on their ascriptions of praise. It is my contention that there is an alternative way of explaining the asymmetry of Knobe's data. On my view, Knobe gets disparate results because his praiseworthy help conditions are not proper analogues for his blameworthy harm conditions.

In order to see why this is the case, we should first look at Knobe's harm conditions. For instance, when we read that the chairman doesn't care that his plans will harm the environment, it is natural – and seemingly justified – for us to form a negative evaluation of the chairman. Similarly, when we read that the lieutenant in the harm condition does not care about putting his soldiers in the line of fire we understandably form a negative evaluation of the lieutenant. In both cases, because the two respective agents do not care about something that it is clear that they should care about we view them in an unfavorable light. This negative evaluation presumably explains subjects'

willingness to a) blame the chairman and the lieutenant in the harm conditions, and b) to say that the negative side effects of their respective actions were brought about intentionally.

Once we take a closer look at Knobe's help conditions, however, we quickly find that the respective agents are not genuinely praiseworthy, as they would need to be in order for the help conditions to be isomorphic with the harm conditions. As Adams and Steadman correctly point out, "the language of the harm conditions seems natural (if uncaring), but the language in the help condition seems highly strained" (2003a). For instance, when we read that the chairman does not care that his plans will help the environment, we are not inclined to view his lack of concern favorably. After all, even the greediest corporate executive would presumably be pleased to find out that a business plan will not only yield a huge profit, but that it will also help the environment. So, the natural reaction – i.e. the reaction we would expect from such a chairman – would be, "Great! You mean to tell me that not only am I going to make large sums of money, but I will also be helping the environment as well? I couldn't have planned this any better! This should keep the EPA off our backs for a while!" And the same thing can be said in the case of the lieutenant who curiously – and seemingly callously – doesn't care that his soldiers will be removed from the line of fire. Once again, we expect a very different response. Rather than not caring at all about the welfare of his soldiers, we expect him to say something like, "Great! You mean that not only will I take Thompson Hill, but in doing so I will actually be saving my soldiers' lives? Things couldn't have turned out any better!"

In both of these cases it is precisely because we find the agent's lack of concern very puzzling and highly inappropriate that we form negative evaluations of them. In this respect the harm conditions and the help conditions are quite similar. Indeed, in all four cases we do not form good opinions of the respective agents because each of them does not care about something that we think they should care about. By my lights, this not only explains why the subjects who received the harm conditions *did* judge that the agents deserved to be blamed for bringing about bad side effects, but it also explains why the subjects who received the help conditions *did not* judge that the agents deserved to be praised for bringing about good side effects.

Thus, Knobe's conclusion – *viz.*, that people's judgments about the intentionality of side effects are affected by negative but not positive moral considerations – only appears to be supported by his data because of the aforementioned problems associated with his vignettes. On my reading, what the results of his side effect experiments really show is that insofar as subjects judge that an *agent* is blameworthy, they are more inclined to say that any *negative* side effects brought about by the agent are intentional and any *positive* side effects brought about by the agent are not intentional. Presumably this is because if the subjects who read one of the help conditions said that the side effects were brought about intentionally, they would have felt compelled to say that the agent deserved praise for intentionally bringing them about – something they were loathe to do insofar as they viewed the agents in the help conditions in a very negative light owing to their blatant lack of concern. But to the extent that even Knobe's *help* conditions involve *morally blameworthy agents*, his data do not rule out the possibility

that if people were presented with a *morally praiseworthy agent*, they would be inclined to judge that the agent brought about the side effects of her action intentionally.

IV: Praise and Side Effects: A New Experiment

To see whether praise could affect people's judgments about the intentionality of the side effects of an agent's actions I conducted another simple experiment. This time the subjects were 40 undergraduates. Each received the following vignette:

Imagine that Steve and Jason are two friends who are competing against one another in an essay competition. Jason decides to help Steve edit his essay. Ellen, a mutual friend, says, "Don't you realize that if you help Steve, you will decrease your own chances of winning the competition?" Jason responds, "I know that helping Steve decreases my chances of winning, but I don't care at all about that. I just want to help my friend!" Sure enough, Steve wins the competition because of Jason's help.⁵

Then each of the subjects received the following two questions: First, did Jason intentionally decrease his own chances of winning the competition by helping Steve?

Second, how much praise does Jason deserve on a scale from 0 to 6 – 0 being no praise and 6 being a lot of praise – for decreasing his chances of winning in order to help his friend? The results were as follows: 55% of the subjects judged that Jason *intentionally* decreased his chances of winning. Moreover, the average praise rating for those who said that he intentionally decreased his chances of winning intentionally was 4.0, whereas the average praise rating for those who said that he did not intentionally decrease his chances was only 2.5. This suggests that there is a positive correlation between how much praise subjects attributed to Jason and whether they judged that the side effect was brought about intentionally.

When we compare the data from my side effect experiment with Knobe's data, we get the following results: whereas 55% of my subjects judged that Jason brought about

⁵ I would like to thank Joshua Knobe for helping with the construction of this vignette.

the side effect intentionally with an average praise rating of $M=3.3$, only 26.5% of the subject in Knobe's CEO and lieutenant cases judged that the side effects were brought about intentionally with an average praise rating of $M=1.4$. I suggest that the best explanation for the asymmetry in the results of our respective side effect experiments is that in my help vignette *Jason's lack of concern is itself morally praiseworthy*. So, when my subjects read that Jason cares more about helping his friend than about winning the competition – an admirable and noble lack of concern – they were presumably inclined to view him in a very favorable light. This explains why the subjects in my experiment were much more likely to judge that Jason deserved praise and thus that the side effects of his actions were brought about intentionally.

These findings notwithstanding, there is something else that is noteworthy about my side effect vignette – *viz.*, even though Jason is morally praiseworthy for not caring about decreasing his chances of winning, this side effect of his action is still a *negative* one. Might we be able to further magnify the effect of praise on subjects' ascriptions of intentional action by matching a praiseworthy agent to a positive, rather than negative, side effect? Originally, I thought this is exactly what would happen. However, when I tried to come up with this type of scenario – while at the same time avoiding fanciful situations involving amnesia, robots, causal deviance, zombies, possible worlds and other common philosophical fare – I quickly discovered an interesting fact about side effects. If an agent is a morally praiseworthy person and the *supposed* side effect she knowingly brings about is morally positive, then it isn't really a side effect for the agent at all. To see that this is so, try to imagine a scenario that has the following three features: a) a morally praiseworthy agent, b) a morally positive action, c) a foreseen yet undesired and

unintended morally positive side effect. On the surface, this does not appear very difficult. But in this case, appearances are misleading to the extent that satisfying the first two conditions seemingly undermines your ability to satisfy the third.

Keep in mind, in order for something to count as a side effect it must be a foreseen yet *undesired* and *unintended* result of an agent's actions. But if the agent is a good person and knows that her performing some action *p* will produce some other positive result *q*, then presumably *q* would simply become an additional part of her reason for *p-ing*. And once an agent desires both *p* and *q* and desires and intends to bring *q* about by *p-ing* – i.e. once *q* ceases to be undesired and unintended – *q* ceases to be a side effect. To see that this is so, let's return briefly to Knobe's second CEO case. As we saw earlier, the problem with this case is that the chairman of the board is blameworthy for *not* caring that the new program will *help* the environment. One easy way to fix this problem would be to simply make the chairman an environmentalist who is happy to hear that the new program will have positive environmental effects. However, once we change the scenario in this manner, helping the environment ceases to be an undesired and unintended side effect of the chairman's actions. After all, insofar as the he is an environmentalist, once the chairman hears that the new program will be beneficial to the environment, he has an *additional reason to adopt the program*. But if helping the environment is one of the chairman's reasons for adopting the new program, it is not really a side effect.

We run into the same difficulty with Knobe's lieutenant case as well. Once again, the problem with the lieutenant is that he does not care about something – *viz.*, the welfare of his soldiers – that he should care about. And once again we cannot simply

correct for his lack of caring by having him care about the welfare of his soldiers without at the same time giving the lieutenant an additional reason for “taking control of Thompson Hill.” Consequently, in trying to change the by attempting to pair a morally praiseworthy agent with a morally praiseworthy side effect, we inadvertently convert the scenario into a non-side effect case. And as far as I can tell, so long as side effects must be foreseen yet undesired and unintended, this problem will seemingly arise no matter how we try to change Knobe’s two original “morally positive” vignettes.

If this is correct – which is to say, if there can be no such thing as a positive side effect so long as the agent in question is praiseworthy – then using examples like Jason’s decreasing his chances of winning may be the only way that we can get at whether positive as well as negative considerations affect folk ascriptions of intentionality in side effect cases. Moreover, this conceptual limitation on side effects may also help explain why the subjects in my praiseworthy side effect experiment were less inclined to judge that Jason intentionally decreased his chancing of winning than the subjects in my praiseworthy non-side effect experiment were to judge that Fred intentionally prevented the explosion. After all, in the latter case Fred is not only praiseworthy, but his actions are morally positive – a combination that is seemingly not possible in side effect cases. In any event, I believe that the preliminary data I have presented here helps to establish that both blame and praise can act expansively on folk ascriptions of intentional actions even in side effect cases – even if blame admittedly has a more pronounced effect. However, before any definitive conclusions about folk ascriptions of intentional action

can be reached, more research must be done. Minimally, I hope this paper serves as yet another small stepping-stone in that direction.⁶

⁶ I would like to thank Joshua Knobe, Al Mele, Virginia Tice, and Rob Woolfolk for helpful comments on earlier drafts of this paper.

References:

- Adams, F. 1986. Intention and intentional action: the simple view. *Mind and Language*, 1: 281 – 301.
- Adams, F. and A. Steadman. 2004a. Intentional action in ordinary language: core concept or pragmatic understanding? *Analysis*, 64:2.
- _____. 2004b. Intentional actions and moral considerations: still pragmatic. *Analysis*, 64:3.
- Audi, R. 1986. Acting for reasons. *Philosophical Review*, 95: 511 – 46.
- Bratman, M. 1984. Two faces of intention. *Philosophical Review*, 93: 375 – 405.
- Butler, R. 1978. Report on *Analysis* “problem” no. 6. *Analysis* 38: 113 – 14.
- Churchland, P. 1981. Eliminative materialism and the prepositional attitudes. *The Journal of Philosophy*, LXXVII, 2: 67-90.
- Davis, W. 1984. A causal theory of intending. *American Philosophical Quarterly*, 21: 43 – 54.
- Duff, R.A. 1990. *Intention, Agency, and Criminal Liability*. Oxford: Basil Blackwell.
- _____. 1982. “Intention, Responsibility, and Double Effect.” *The Philosophical Quarterly*, Vol. 32, No. 126: 1 – 16.
- Finnis, J. 1991. “Intention and Side-Effects.” In R.G. Frey & C. Morris, eds. *Liability and Responsibility*. Cambridge: Cambridge University Press.
- Gore, M. 1996. “Should the Law Distinguish Between Intention and (Mere) Foresight?” *Legal Theory*, 2: 359 – 380.
- Harman, G. 1976. Practical Reasoning. *Review of Metaphysics*, 79: 431 – 63.
- Hart, H.L.A. 1968. *Punishment and Responsibility*. Oxford: Oxford University Press.
- Katz, L. 1987. *Bad Acts and Guilty Minds*. Chicago: University of Chicago Press.
- Knobe, J. 2003a. Intentional action and side effects in ordinary language. *Analysis* 63: 190 – 94.
- _____. 2003b. Intentional action in folk psychology: an experimental investigation. *Philosophical Psychology* Vol.16, No. 2: 309 – 24.
- _____. 2004. Intention, intentional action, and moral considerations. *Analysis*.
- Lyons, R. 1976. “Intention and Foresight in Law.” *Mind*, JA 76; 85: 84 – 9
- McCann, H. 1986. Rationality and the range of intention. *Midwest Studies in*

- Philosophy*, 10: 191 – 211.
- _____. 1991. Settled objectives and rational constraints. *American Philosophical Quarterly*, 28: 24 – 36.
- Mele, A. 2001. Acting intentionally: probing folk notions. In *Intentions and Intentionality*, eds. B.F. Malle, L.J. Moses, and D.A. Baldwin, 27 – 43. Cambridge: MIT Press.
- Mele, A. and P. Moser. 1994. Intentional action. *Noûs* 28: 39 – 68.
- Mele, A. and S. Sverdlik. 1996. Intention, intentional action, and moral Responsibility. *Philosophical Studies* 82: 265 – 87.
- Nadelhoffer. T. 2004. “The Butler Problem Revisited.” *Analysis* 64: 277 – 284.
- _____. Forthcoming. Skill, Luck, and Folk Ascriptions of Intentional Action. *Philosophical Psychology*.
- Oberdiek, H. 1972. “Intention and Foresight in Criminal Law.” *Mind*, JL 72, 81: 389 – 400.
- O’Shaughnessy. B. 1973. Trying (as the mental “pineal gland”). *Journal of Philosophy*, 70: 365 – 86.