

Folk Concepts, Surveys and Intentional Action

Frederick Adams & Annie Steadman

University of Delaware

“...if there is a widely shared concept of intentional action...a philosophical analysis of intentional action that is wholly unconstrained by that concept runs the risk of having nothing more than a philosophical fiction as its subject matter” (Mele, 2001, p. 27).

1. Introduction

In a recent paper, Al Mele (2003) suggests that the Simple View of intentional action is “fiction” because it is “wholly unconstrained” by a widely shared (folk) concept of intentional action. The Simple View (Adams, 1986, McCann, 1986) states that an action is intentional only if intended. As evidence that the Simple View is not in accord with the folk notion of intentional action, Mele appeals to recent surveys of folk judgments by Joshua Knobe (2003, 2004a, 2004b). Knobe’s surveys appear to show that the folk judge unintended but known side effects of actions to be performed intentionally. In this paper we will reject Mele’s suggestion that the Simple View is “fiction.” We will also discuss the relationship between surveys and philosophical theories, and the abilities of surveys to access folk core concepts. We will argue that considerations of both fail to support Mele’s suggestion.

2. Knobe’s Surveys

In a series of experiments, Joshua Knobe (2003b) asked subjects questions about intentional actions. 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' (4). 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 that one can intentionally do something without intending it, while in the help condition, the folk judgments seem to accord with the Simple View. This difference in judgments needs to be explained.

Knobe conducted a second experiment in order to validate his results. Knobe explained why he ran the second experiment, saying: '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' (4). The second experiment was structurally identical to the first, but instead the vignettes are about sending soldiers to their possible doom. 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 were highly statistically significant.

3. Knobe's explanation of his data

Initially, Knobe (2003b, p. 7) 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.” 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.

In another paper, Knobe (2003a) says: “evaluative considerations do play some role in people's concept of intentional action.” He believes his data show folk core concepts of intentional action entail that actions can be intentional, though not intended. They would be deemed intentional because of moral considerations. 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 in one's folk concept for whether the action is intended. If true, this would indeed be surprising.

Knobe (2003a) calls his hypothesis a “dissociation” hypothesis. Normally, intentionally doing an action is associated with intending it and being skillful at doing it. One who is good at playing darts may be deemed to have intentionally thrown a bull's eye when wanting to, intending to, and being skillful at doing so. Knobe looks at a series of examples which vary the degree of skill of an actor. As the skill diminishes, fewer people tend to say the action was performed intentionally. The one exception is in cases where there are moral considerations. If one lacks skill in shooting a gun and shoots at a rival from a great distance, people still judge the

success as intentional shooting. Moral considerations override the usual association of skill and intentional outcome. Similarly, Knobe would say that there is a dissociation of the intention from intentional action, in cases where there are moral considerations about the outcome. The case of the CEO harming the environment is a paradigm case of dissociation. The CEO is deemed to have harmed the environment intentionally, though he did not intend the harm.

We are inclined to agree that moral considerations cause subjects to dissociate intentional actions from skill and intention. For us the question is why? Is it because there is a fully articulated folk core concept of intentional action that includes a semantic marker for moral consideration? Or is there another explanation? We believe there is another, better explanation of the dissociation phenomenon. We believe that there are pragmatic factors influencing folk judgments. We do not think these pragmatic factors are part of a fully articulated folk core concept of intentional action, as we shall explain.

4. Surveys and folk core concepts

We find Knobe's surveys to be important and interesting, as well as surprising. In particular, we find surprising the asymmetry of judgment in the "help" and "harm" conditions. We agree that he is accessing something in people's judgments, but we doubt that he is accessing fully articulated core concepts. 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 factor 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. For instance, if an intention is connected by causal deviance to its conditions of satisfaction, the action is not done intentionally. 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.

Another factor involves skill (Mele 2001, Mele & Moser, 1994). Is it possible for a non-skilled player to intentionally make a basket from the half-court line? Even among philosophers, there is a difference of opinion about not only how skilled one must be to make such a shot intentionally, but how confident one must be. Paul Grice (1971) claims one must believe that one will make the basket in order to intend and, ultimately, do it intentionally. Robert Audi (1973) believes that one's chances must be at least as good at making it as not making it in order to intend to make the shot. Anything less and one is merely trying with the hope of success. Myles Brand (1984) believes that even with "serious doubts" about success, one may intend to do something (such as make the basket), and Hugh McCann (1998) maintains that any goal that guides one's planning and acting behavior is one that is intended (regardless of how low one's subjective estimates of success may be).

These philosophers are paid to have articulated concepts of intentional action and they do not agree. Surely it is unlikely that the folk share a single articulated concept of intentional action, in light of this.

Even the folk are not consistent as to what constitutes the elements of intentional action. 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.¹

A possible reason for this is that ordinary life seldom requires one to have a theory of intentional action finely articulated to the point of being able to differentiate acting knowingly from acting intentionally. In some circumstances, whether one's act is intentional or not is irrelevant. Using mouthwash kills bacteria. No one worries about whether or not they are killed intentionally. Driving one's car wears down the rubber on the tires. No one thinks about whether this is done intentionally (in most circumstances). Of course, in the right context, both of these could have significance, but that is not the norm. In other circumstances, it is obvious that people act on purpose, and the folk don't need to articulate the cognitive or motivational conditions on the nature of intentions. If one robs a bank, we know this was done on purpose or intentionally, even if we don't know what it takes to intend to rob the bank. Did one merely hope to rob it, though she was not confident that she would succeed? These sorts of considerations seem not to apply. We realize that in the law, finer distinctions are made (such as first degree murder, second degree murder, and manslaughter). The average person only makes these distinctions in the courtroom setting.² So the "folk" concept is not necessarily in synch with legal concepts of "intent."

We have now given some reasons to think that the folk are not likely to have fully articulated concepts of intention and intentional action. We will now give alternative interpretations of Knobe's survey data. Later, we will report data from our own surveys which tend to support our interpretation of Knobe's data.

¹ 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.

² We discuss these legal distinctions further in Section 6 below.

4a: Our interpretation: It's Pragmatics

Even if the folk do not have a fully articulated core concept of intentional action, 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' or 'you did that intentionally, didn't you?' are social ways to assign blame and of discouraging actions that one disapproves. 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.

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.³ However, the folk may associate intentionality with judgments of praise and blame due to social or evolutionary pressure (Cosmides & Tooby, 1994, Mithen, 1996).⁴ 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.

If the chairman in Knobe's vignette literally 'does not care at all about the environment' then he does not intend to help or harm it. Yet in the harm condition, folk largely judge that the chairman's harming the environment was intentional. We suspect that what is going on in the

³ A morally neutral act such as setting one's watch can be perfectly intentional and yet be worthy of neither praise nor blame.

⁴ Consider Cosmides & Tooby's 'cheater-detection' modules which work on purposive behavior, or Mithen's claim that the modern human mind evolved for the purpose of tacking social facts.

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.⁵ When subjects deem the chairman's act is intentional in the harm condition, they are more likely accessing the pragmatic features of the intentional talk, than accessing an articulated core concept.

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

⁵ We conducted surveys in which we test to see whether subjects display cognitive inconsistency when confronted with it. The results will be discussed at length below.

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 (not to mention downright cold). 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.

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 the view that intentional actions are intended. So in the help condition, people may be making the right judgment, but for the wrong (pragmatic) reason. Or, due 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. 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 the view that intentional

actions are intended, but mainly because no clearly articulated core concept of intention or intentional action is being consulted.

In his first study, 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 (Malle & Knobe, 1997, p.116).

The subjects may blame in the harm condition on the basis of their disapproval of the attitudes 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 the view that intentional actions are intended, it offers another way of explaining the asymmetry of Knobe's data without abandoning the view that intentional actions are intended.

4b: McCann's Interpretation

McCann (forthcoming) noted that Knobe did not ask subjects in the chairman vignette whether or not the chairman intended to harm the environment.⁶ Later Knobe (2004) did ask some subjects in a follow-up experiment whether the chairman intended to harm the environment, and asked others whether the chairman harmed the environment intentionally. However, McCann

⁶ We too (Adams & Steadman, 2004a) worried about the fact that Knobe did not survey to see if the "folk" were attributing intentions to the CEO.

remained unsatisfied that Knobe did not ask the *same* subjects *both* questions. The thinking was that the same subjects who said the chairman harmed the environment intentionally might be the very ones who would attribute to the chairman the intention to harm the environment (if asked). McCann (forthcoming, p. 7) surmised that, to be consistent, subjects would say yes or no to both questions, speculating "... if respondents had to face the alleged contradiction directly [of judging one to have acted intentionally without intending it], they would be more inclined to treat intentionality and intention as going together...." So with Knobe's help McCann devised a survey of his own, concentrating exclusively on the "harm condition." They divided the pool of subjects into two groups. One group would receive both questions, and the other group would receive only one question (either the question "Did the chairman intend to harm the environment?" or "Did the chairman harm the environment intentionally?") McCann surveyed 106 subjects. Of those asked one question, 63% said the chairman's act of harming the environment was intentional. 27% said the chairman intended to harm the environment. Of those asked both questions, 80% said the chairman's act of harming the environment was intentional. Only 12% said that the chairman intended to harm the environment. According to McCann, the results for those asked both questions were "completely unanticipated" (p. 9).

McCann and Knobe repeated a slightly modified experiment, but found strongly similar results. In the modified experiment, they asked not whether the chairman *had the intention* of harming the environment (which suggests that he had only one main intention), but whether he *intended* to harm the environment (suggesting that he may have had that intention, among others). In this case there were 99 subjects. Of those who got one question only, 64% said the chairman's action was intentional, and 42% said his action was intended. Yet for those asked

both questions, 75% said the chairman's action was intentional and only 31% said it was intended.

McCann, a supporter of the view that intentional actions must be intended, took solace in the result that subjects are "more willing to say the chairman "*intended* to harm the environment than that it was his intention to do so" (10). He took solace in this because he thinks that change in phrasing of the questions has a significant impact on the subjects' responses. When asked about "the intention" of the chairman, fewer attribute to him the intention to harm the environment, than when asked about "an" intention (even though this is not the exact wording McCann and Knobe chose). Although this does represent a significant change, the asymmetry in the condition where subjects were asked both questions remains stunning. As McCann says, subjects do not appear to "shy away" from any contradiction implicit in their verdicts (as would be true, if their folk concept of intentional action required that it be intended). In the end, McCann does not offer a convincing explanation of this asymmetry in folk judgment. He does say the results one gets in such surveys depends upon the way the questions are worded and that different wordings may not be correctly accessing what the folk really think (for a variety of reasons).

For our part, we are happy that McCann does acknowledge that in these survey situations there is something that "has partly to do with the pragmatics of the situation—with the fact that to deny the chairman's action was intentional is to suggest it was not blameworthy....then factors that appear to tip the results...may for the most part be pragmatic, rather than anything to do with the semantics of the terms 'intentional' and 'intend'" (p. 13).

Yet, McCann thinks this still doesn't explain well the asymmetry in the harm and help conditions. It doesn't explain why the CEO was perceived as blameworthy. McCann does not

think that it can be as simple as that the CEO is perceived to knowingly (but not intentionally) have done something wrong. The CEO is not praised for knowingly helping the environment in the help condition. So McCann thinks there is more at stake.⁷

McCann wants to say that, for the harm condition, the folk are judging the *act itself* (15). They believe there is a “perfect duty” not to harm the environment that cannot be overridden or set aside for profits. McCann adds that, since the CEO goes forward anyway, “there is ...an additional dimension of intending in the harm vignette. In setting aside the requirements of morality the chairman resolves—that is, forms the intention—to put his own and his company’s projects above moral principle, and in moving ahead he intends to do exactly that” (16).

For McCann this means there is a “hidden variable” in the vignette...an extra intention. The CEO *does intend* to harm the environment, after all. But why, then, don’t the subjects see fit to say yes to the “intend/intention” questions, if they see fit to say yes to the “intentional” questions? If the Simple View that intentional acts are intended is what is prompting these results in the harm condition, as McCann suggest, then why doesn’t the same mechanism bring the exact same numbers of responses for both questions, when both are given to the same subjects?

First, McCann explains this away by saying “the intention to harm the environment is not part of the *content* of the chair’s decision in the harm vignette” (p. 16). True, but remember this second intention is “hidden” (not explicitly mentioned in the wording of the vignette). One need not find it in the vignette itself, on McCann’s view. Hidden variables still have causal influence on his view, so that can’t explain why subjects don’t answer both questions the same.

⁷ We don’t think there is more at stake because we attribute lack of praise to the CEO’s admission that “he doesn’t care at all about helping the environment.” Subjects do not wish to praise such an attitude, even if it leads to helping the environment knowingly.

Second, McCann renews his claim above that “we are all taught as children that when you have knowingly done wrong, you don’t get to not mean it. Just the opposite: you did wrong on purpose” (17). So, again, the CEO does intend the harm after all. But if subjects see the hidden variable (see the hidden intention), why don’t they show it by marking the questions about intent and intentional action of the CEO the same? McCann has no good explanation of this.

Now, again, we think the lesson is not that if one does wrong knowingly one *intended* to do wrong, but rather that if one does wrong knowingly, one is blameworthy. Being blameworthy, the best way to make it stick is to say “you did that on purpose,” or “you did that intentionally.” And since, as we will see below, people can make the “knowingly” vs. “intentionally” distinction in general, we maintain that there is in addition some sort of socialized *pragmatic override* of the semantics going on in these morally charged cases. Since McCann acknowledges the possibility of something pragmatic going on, our main disagreement is over just how much is attributed to the pragmatics.

McCann thinks the folk judgments are consistent with the view that intentional actions are intended because the folk are judging the character of the actors in light of their actions themselves. We see this point as consistent with our claim that it is the pragmatics of praise and blame that is at work in Knobe’s vignettes and the surveys. The difference between our account and McCann’s is that we give a mechanism—the pragmatics of intentional language and judgments—to explain the phenomena.

McCann explains the asymmetry of folk judgment by giving up on the distinction between acting knowingly and intentionally *only when the known consequences are bad* (harm condition). We think this is unprincipled, to put it mildly. On his view, you have result that the folk blur the

judgment that something is done wrong knowingly with the judgment that something is done wrong intentionally. This appears to be McCann's view of the folk and part of his explanation of why they deem the chairman's acts to be harming the environment intentionally. In what follows, we will present data that show the folk *do make the discrimination* between acting knowingly and acting intentionally. So the asymmetry of judgment in the harm vs. help conditions cannot be explained solely in these terms. McCann too needs some explanation to play the role of our *pragmatic* override in the harm condition. Merely saying one does "not get not to own it," doesn't fully explain. It doesn't distinguish between the CEO being blameworthy because he caused harm knowingly and the CEO being blameworthy because he caused harm intentionally. In both cases, *you don't get not to own it*. Thus, McCann's explanation is still incomplete.

5: Our new data

In previous papers (Adams & Steadman 2004a, 2004b), we maintained that Knobe's surveys did not give respondents an opportunity to distinguish an act's being done knowingly from its being done intentionally. We suggested that if given the option of saying the chairman harmed the environment knowingly versus harmed the environment intentionally, a significant number would opt for the former. We thought this was important because in Knobe's surveys, he gave the folk only the option of saying the chairman's act was intentional or not intentional. Since subjects were inclined to accept that it was not the chairman's intention to harm the environment, their judging that the chairman harmed the environment intentionally (though the only option given by Knobe) may actually create cognitive tension. We surmised that subjects would opt for judging that the chairman harmed the environment knowingly, but not intentionally, if given the option. This choice would relieve that cognitive tension, and, at the same time, allow subjects to

blame the chairman for knowingly harming the environment. We believe that a large number of the respondents in Knobe's original experiments only said the chairman harmed the environment "intentionally" because they wanted to blame his actions.

We maintain that acting knowingly versus intentionally is a real distinction that most folk know how to make, and do make. A simple example is one such as the following. When we drive our cars, we wear down the tires and know that we do. However, it is no desire of ours nor plan that we act upon to wear down the tires. We do so knowingly, but not intentionally. Were we to do so intentionally we might squeal the tires, engage in fast starts and stops, corner hard, etc.

A more exotic example is the following. Trapped on a narrow ledge with no way to escape but to jump over a chasm, Al might decide to try to jump to escape. Al knows that his chances of success are slim and knows that his jumping may ultimately cause his death. If Al jumps and does not make it, he causes his death knowingly, but he does not commit suicide (i.e., intentionally cause his own death). We were convinced that the folk can and do make this distinction. So we designed a survey of our own to test this.

With the above considerations in mind, we designed and conducted a new experiment. The goals of the study were to A) see what would happen if subjects were offered the opportunity to distinguish an action performed "knowingly" from an action performed "intentionally" in their responses, and B) to test subjects' usage of intentional language in a morally neutral scenario (thus removing the praise/blame pragmatic implication issues from the equation.) We wanted to remove the praise/blame aspect of the judgment because it is our hypothesis that the desire to blame the chairman for his morally bad actions can interfere with

one's ability to discriminate actions done knowingly from those done intentionally. After all, they will judge an action to be intentional even if not intended, when there is blame to be levied.

We surveyed 49 Summer Undergraduate Researchers at the University of Delaware. The subjects were randomly divided into two groups.

The first group received Knobe's "harm" vignette, identically reproduced. They were then asked to answer the following question:

Please select the best answer:

_____ "The chairman harmed the environment knowingly, but not intentionally."

_____ "The chairman harmed the environment knowingly and intentionally."

20% of respondents answered "knowingly, but not intentionally," indicating that the folk are capable of making such a distinction when offered the chance. Still, 80% of subjects said that the action was done "knowingly and intentionally," telling us that their desire to blame is hard to override. These results demonstrate that the folk *can* and *do make* the distinction between acting "knowingly" and "intentionally," as well as showing just how difficult it is to override the folk's pragmatic programming to use intentional language to levy blame.

The second group received the same vignette with the following modifications designed to make the scenario morally neutral while maintaining Knobe's structure:

The vice-president of a company went to the chairman of the board and said, "We are thinking about starting a new program. It will help us increase profits, but starting the new program will tip off our competitors."

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

They started the new program. Sure enough, this tipped off their competitors.

The subjects were then asked the following question:

Please select the best answer:

_____ “The chairman tipped off the competitors knowingly, but not intentionally.”

_____ “The chairman tipped off the competitors knowingly and intentionally.”

For this question, 71% of respondents chose “knowingly, but not intentionally,” whereas only 29% went for “knowingly and intentionally.” Note that the results in this morally neutral scenario are exactly the opposite of what one receives in the “harm” condition.

A summary of our data is as follows:

	Environment “harm” vignette	Tipping off the competition Vignette
Knowingly and intentionally	80%	29%
Knowingly, but not intentionally	20%	71%

In all of these vignettes, the chairman clearly states that he “does not care at all” about the suggested side-effect. If one truthfully does not care at all about a certain possible outcome, it makes no sense to say that he or she intended it. Even in Knobe’s experiments, few of the folk attribute the intention to harm the environment to the chairman. Our results demonstrate that the folk are capable of making the distinction between a result caused intentionally and a result caused knowingly. It is only in a morally neutral case, however, that the majority of respondents are freed of their pragmatic programming so that they can make judgments purely on what they

think about intentional action; and it seems that here their thinking is in keeping with the view that intentional actions are intended.

We attribute the difference between the high number of subjects who pick “knowingly and intentionally” in the “harm” case versus the high number who pick “knowingly but not intentionally” in the “morally neutral” case to pragmatic override. That is, we believe that even if one can make the distinction between acting knowingly and acting intentionally (as our data indicate that they clearly can), their life-long habit and social training of pragmatic use of intentional language to assign blame is hard, if not impossible, to override. Both Mele and McCann much earlier had suggested to Knobe that subjects might be instructed that a person can be blameworthy for an act that one did not do intentionally. And Mele and McCann both believed that this “instruction” might cause subjects to respond differently in the “harm” kinds of cases. It was even suggested that instead of “instructing” subjects, one might draw them to this realization by a type of “Socratic questioning.” However, Knobe (2003a) conducted experiments in which the “instruction” had no significant effect on the responses in “harm” conditions of his surveys. This shows how deeply ingrained is the pragmatic use of intentional language, even when it has been recently brought to subject’s attention that blame does not require attribution of intention. Attribution of intention is the usual way, the habitual way of assigning blame—a habit that is very hard to break or override. We suspect that part of the explanation for this is that subjects compartmentalize. In one cognitive compartment they keep the information that it is indeed possible to blame without attributing intention or intentionality. Yet, when the situation for blaming arises, from another cognitive module, their tendency to blame kicks in and overrides information stored in the other module.

Compartmentalization is not an uncommon phenomenon. For example, in Introduction to Philosophy, students learn to critique the arguments for the existence of God. On an exam where they give these critiques of the arguments, if asked “Do you believe that God created the world?” very many will say “Yes, the world had to have come from somewhere.” This answer discounts their own critique of the Cosmological Argument earlier in the same exam. Even if they appear to find the critique of the Cosmological Argument convincing, many students tend to compartmentalize and not let their critique interact with their religious convictions formed prior to taking the class.

6: Surveys and philosophy

The challenge from the opening quote by Mele seems to suggest that one’s philosophical theory of action should conform to folk judgments (concepts) of action. If they don’t, the philosophy is fiction. We see several problems. First, we deny that the folk have clearly articulated concepts or theories of intentional action. We have given reasons above to believe that even philosophers disagree about these matters, when they try to articulate such concepts (theories). Second, why think that a philosophical theory should conform to folk judgments? We wouldn’t think that relativity physics, or mathematics of orders of infinity, should conform to folk judgments. So why should philosophical theories of action? The only answer we can think of is that Mele must not think philosophical theories go that deep. The nature of intentional action is not hidden (in the way the fundamental nature of reality might be) from the reasoning of the common man. He seems to think that whatever intentional action is, the folk ought to be able to figure it out. What is surprising to us is that Mele has spent his philosophical career working on articulating concepts and theories about the nature of intentional action. It can’t be so easy that the folk all

do it. Third, we suspect that the correct role for surveys of folk judgments is to measure what people think about reality (intentional action) not what reality (intentional action) is.

We agree that there can be philosophical interest in finding out what the folk think about intentional action, and why they think it. That's why we've used surveys ourselves.

Nonetheless, we have a deeper concern whether or not surveys can access the kinds of subtle distinctions we are appealing to in our explanations of folk judgments. That is, we are not sure that either Knobe's or our own surveys are sensitive enough to determine whether subjects are making judgments based upon the semantics of intentional language or the pragmatic implications of the use of that language. We suspect that none of the surveys we've seen up to now are able to make these discriminations, and we have not yet decided how to go about testing for them in this context. We've tried to begin to get at such differences by seeing if subjects can distinguish doing something knowingly from doing it intentionally. Still, how can a subject's checking a box on a survey tell us whether it is the semantics or the pragmatics of the terms they are employing that is responsible for the answers? Until we have an answer to this question, we don't think the dispute between us and Knobe, Mele, McCann, and others will be able to be decided by surveys. We do think that it is the pragmatics of intentional language that is responsible for the results that Knobe, McCann, and others are getting on the surveys. We have given philosophical reasons why think this. But we don't think surveys alone can decide (at least, not those employed so far).

While we have argued that the folk do not have a fully articulated concept of intentional action, we do believe that the seeds of the distinction we test for between acting intentionally and knowingly do exist in the minds of the folk. This is of practical importance for the Law. Thus, our philosophical view of intentional action is not going to be "fiction" by Mele's lights. We say

this because we find reflected in the law this very distinction between acting knowingly and intentionally. In the Delaware State Code, Chapter Two, Section 231, the following distinctions are made concerning state of mind of agents committing offenses:

(a) “Intentionally”, -- A person acts intentionally with respect to an element of an offense when:

(1) If the element involves the nature of the person’s conduct or a result thereof, it is the person’s conscious object to engage in conduct of that nature or to cause that result; and

(2) If the element involves the attendant circumstances, the person is aware of the existence of such circumstances or believes or hopes that they exist.

(b) “Knowingly”, -- A person acts knowingly with respect to an element of an offense when:

(1) If the element involves the nature of the person’s conduct or the attendant circumstances, the person is aware that the conduct is of that nature or that such circumstances exist; and

(2) If the element involves a result of the person’s conduct, the person is aware that it is practically certain that the conduct will cause that result.

The code goes on to specify even more differences in state of mind for acts done “recklessly,” out of “criminal negligence,” or simply “negligence,” as well.

Concentrating mainly on the distinction between acting intentionally and knowingly in the law, we see that the code identifies an intentional act with one where the agent has the act consciously as a goal object, and believes or hopes that the circumstances necessary for carrying out the action exist. Whereas, the agent who acts knowingly is aware that his conduct is of such a nature that it may bring about the result and is aware that it is practically certain that his conduct will cause the result. What is missing in the latter case is the conscious goal.

We would further point out that the distinction is clearly written to categorize acts that may be illegal. This shows that at least the law can override the pragmatics of intentional talk and realizes that some things may be blameworthy and punishable, even when done “knowingly” but not “intentionally.”

7: Conclusion

In this paper, we have offered a defense of the Simple View of intentional action in the face of Mele's claim that it is "fiction," if not in accord with empirical folk surveys such as those first used by Knobe. We have offered an alternative explanation of Knobe's results (that it is the pragmatics of intentional language) that is consistent with the truth of the Simple View that intentional actions are intended. We've replied to Knobe's objections to our alternative explanation of his results. We considered McCann's further experimental results and his own explanation of them. We've given reasons to think McCann's explanation is insufficient to explain the asymmetry of folk judgments in the "help" and "harm" conditions. We have offered results of our own surveys that support our explanation that it is the pragmatics of intentional language that is accounting for the asymmetry. We discovered that the "folk" can and do distinguish acting knowingly from intentionally. We suggested that they do this because they feel the cognitive pull of the Simple View (that intentional actions are intended). Still we recognize that we need to appeal to mechanisms such as *pragmatic override* and *cognitive compartmentalization* to explain why the folk abandon the "knowingly/intentionally" distinction, when it comes to the harm condition. If there is further empirical work to be done in this area, we think it will be in trying to figure out the cognitive mechanisms of *compartmentalization* and *pragmatic override*.⁸

References

- Adams, F. (1986). "Intention and Intentional Action: The Simple View," *Mind & Language*, 1, 281-301.
- Adams, F. & Steadman, A. 2004a: Intentional Action in Ordinary Language: Core Concept or Pragmatic Understanding? *Analysis*, 64, 173-181.

⁸ We would like to thank the University of Delaware's Office of Undergraduate Research for support for this project. We also thank Al Mele, Joshua Knobe, Hugh McCann, and Shaun Nichols for helpful discussions.

Adams, F. & Steadman, A. 2004b: Intentional Action and Moral Considerations: Still Pragmatic, *Analysis*, 64, 268-276.

Audi, R. (1973). "Intending," *Journal of Philosophy*, 70, 387-403.

Brand, M. (1984). *Intending and Acting: Toward a Naturalized Action Theory*. Cambridge, MA: MIT Press.

Cosmides, L. & Tooby, J. 1994. *The Adapted Mind*. Oxford: Oxford University Press.

Grice, H. (1971). "Intention and Uncertainty," *Proceedings of the British Academy*, 57; London: Oxford University Press.

Knobe, J. (2003a). "[Intentional Action in Folk Psychology: An Experimental Investigation](#)," *Philosophical Psychology*, 16, 309-324.

Knobe, J. (2003b). "Intentional Action and Side Effects in Ordinary Language, *Analysis*, 63, 190-193.

Knobe, J. (2004). "Intention, Intentional Action and Moral Considerations," *Analysis*, 64, 191-187.

Malle, B. & Knobe, J. 1997. The Folk Concept of Intentionality. *Journal of Experimental Social Psychology* 33: 101-121.

McCann, H. forthcoming: "Intentional Action and Intending: Recent Empirical Studies" *Philosophical Psychology*.

McCann, H. (1986). "Rationality and the Range of Intention," *Midwest Studies in Philosophy*, 10, 191-211.

McCann, H. (1998). *The Works of Agency: On Human Action, Will, and Freedom*. Ithaca: Cornell University Press.

Mithen, Steven. (1996) *The Prehistory of the Mind*. London: Thames and Hudson.

Mele, A. (2001). "Acting Intentionally: Probing Folk Notions," in B. Malle, L. Moses, and D. Baldwin, eds. *Intentions and Intentionality: Foundations of Social Cognition*. Cambridge, MA: MIT/Bradford.

Mele, A. (2003). "Intentional Action: Controversies, Data, and Core Hypotheses," *Philosophical Psychology*, 16, 325-340.