

The Folk Concepts of Intention and Intentional Action: A Cross-Cultural Study¹

Joshua Knobe and Arudra Burra

Princeton University

ABSTRACT

Recent studies point to a surprising divergence between people's use of the concept of *intention* and their use of the concept of *acting intentionally*. It seems that people's application of the concept of intention is determined by their beliefs about the agent's psychological states whereas their use of the concept of acting intentionally is determined at least in part by their beliefs about the moral status of the behavior itself (i.e., by their beliefs about whether the behavior is morally good or morally bad). These findings raise a number of difficult questions about the relationship between the concept of intention and the concept of acting intentionally. The present paper addresses those questions using a variety of different methods, including conceptual analysis, psychological experimentation, and an examination of people's use of certain expressions in other languages.

Consider the concept *intention*. This is the concept of a particular type of mental state. We use this concept to predict and explain human behavior. Many cognitive scientists believe that it forms a part of a proto-scientific theory of the human mind.

Now consider the concept *acting intentionally*. It seems that this concept is used to divide behaviors into two basic classes — those that are performed ‘intentionally’ and those that are performed ‘unintentionally.’ But it has proven notoriously difficult to say precisely how this concept works and what role it plays in our lives.

One obvious hypothesis would be that the concept of acting intentionally is closely tied to the concept of intention. In essence, the idea would be that a behavior counts as having been performed ‘intentionally’ if it stands in the right sort of relation to the agent’s intentions. Thus, suppose that the agent insults his neighbor. One might think

¹ We are grateful to Gilbert Harman, Alfred Mele, Daniel Rothschild, Paulo Sousa and Jonathan Weinberg for their very helpful comments on an earlier draft.

that he insulted his neighbor *intentionally* if and only if he had an *intention* to insult his neighbor and this intention led in the right way to his behavior.

As we shall see, experimental research has not been kind to this hypothesis. Systematic studies do not seem to show any simple correspondence between people's use of the word 'intentionally' and their use of the word 'intention.' In fact, it appears that people's use of 'intentionally' does not simply reflect their beliefs about the psychological states of the agent. Rather, it seems that people's use of 'intentionally' is sensitive in a complex way to the *moral* status of the agent's behavior. That is to say, people's intuitions as to whether a given behavior was performed intentionally depend in part on whether they take the behavior itself to be morally good or morally bad.

In light of these findings, it seems worthwhile to consider an alternative hypothesis. Perhaps the concept of acting intentionally is radically unlike the concept of intention. We said above that the concept of intention functions to facilitate predictions of behavior. But perhaps the concept of acting intentionally does not work like that; perhaps it should be understood primarily as a tool for making judgments about whether people deserve moral praise or blame for their behavior.

Our aim here is to investigate this alternative hypothesis using a variety of methods. We review empirical evidence about the use of the two concepts, explore a number of philosophical analyses and — departing somewhat from the usual approach to these issues — report the results of an empirical study of the use of certain words in other languages.

I

Historically, there have been two basic reasons for concluding that the concept of acting intentionally was closely tied to the concept of intention. The first is the alleged relation between the *use* of the two concepts. The second — not often mentioned explicitly, but highly influential nonetheless — is the relation between the *words* used to express these concepts in English. Since the word ‘intentionally’ is so obviously related to the word ‘intention,’ it may be assumed that the concepts expressed by these words are related as well.

This assumption comes out most clearly in J. L. Austin’s (1979) famous discussion of ‘intention’ and ‘intentionally’:

[A]ny unit of speech *U* should sound *tanto quanto* like every other unit of speech that ‘means’ anything like what *U* means, and unlike *tanto quanto* every other unit of speech that means anything unlike what *U* means or that small variations in meaning should be signified by small concomitant variations in sound.

More often, however, the assumption is entirely implicit. It is simply taken as a given that, since the English word ‘intentionally’ sounds so much like the English word ‘intention,’ there must be a tight connection between the concepts these words express.

Thus, Bratman (1985, p. 213) writes:

A theory of intentional action cannot, however, stand alone. It needs to be related to a plausible conception of future intention — intending (or, having an intention) now to do something later. After all, both phenomena in some sense involve *intent*; our theory needs to say in what sense.

Why exactly should we suppose that a theory of intentional action needs to be related to a conception of future intention? Why should we assume that the notion of intent is

involved in both phenomena? Clearly, the chief reason is that the English words ‘intentionally’ and ‘intention’ are both morphologically derived from the English word ‘intent.’

But if the argument is to rest on the morphological relations between specific words, why should discussion of these issues be so completely dominated by words in *English*? Clearly, the reason is not that the English language offers us some unique kind of insight into the fundamental concepts underlying folk psychology. Rather, it just happens to be the case that discussion of these issues has been carried out primarily in English and that participants in that discussion have therefore been influenced more by English morphology than by the morphologies of other languages.

Yet it can hardly be denied that other languages are relevant to the issue at hand. Suppose for a moment that many different languages have a word that functions much as the word ‘intentionally’ does in English. If we find that in all of these languages the word for ‘intentionally’ is morphologically related to the word for ‘intention,’ we would have strong evidence for the view that the corresponding concepts are related in some fundamental way. On the other hand, suppose we find one language in which the word for ‘intentionally’ is morphologically related to the word for ‘wanting,’ another in which it is related to the word for ‘knowing,’ another in which it is related to the word for ‘trying,’ and so forth. Then, surely, we would have little reason to attach any special significance to the fact that English is a language in which the word for ‘intentionally’ is morphologically related to the word for ‘intention.’

Our strategy, then, will be to investigate people’s use of certain expressions in other languages. But first we need to put in place a conceptual framework, and for that, it

will be necessary to look more closely at the use of certain terms in English and at the philosophical analyses that have been proposed to understand these terms.

II

Warning: This next section simply summarizes the key findings from Knobe (2003) and McCann (forthcoming). Readers who are already familiar with those two papers should skip ahead to section III.

We begin by presenting empirical evidence concerning people's use of the terms 'intention' and 'intentionally.' Our key claim here will be that people are sometimes willing to say that a behavior was performed 'intentionally' even in cases where they are not willing to say that the agent had an 'intention' to perform that behavior. This claim does not originate with us — it goes back to the groundbreaking work of Harman (1976) and Bratman (1984, 1987) — but we will be pursuing a slightly different account of the phenomenon here.

Working within English, one sees a clear distinction between *intention* and *foresight*. As one example, consider a corporate executive who decides to implement a new policy. She is aware that the policy will increase profits and that it will also produce some other effect x . But the executive does not care at all about effect x ; all she wants to do is increase profits. Here the executive *foresees* that she will bring about effect x but does not have an *intention* to bring about effect x .

A question now arises as to how the concept of acting intentionally relates to the concepts of intention and foresight. Recent experimental evidence points to a surprisingly

complex relationship among these three concepts. Specifically, it appears that different behaviors differ from each other, with foresight being felt to be relevant to morally bad behaviors in a way that it is not relevant to morally good behaviors. Thus, there can be no single answer as to whether the corporate executive intentionally brought about ‘some effect x .’ The only possible answer is that it depends on whether effect x happens to be morally good or morally bad.

For a simple example, let us introduce the story that we will call the *harm 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.

And now let us contrast this vignette with another — the *help vignette* — that is constructed by replacing the word ‘harm’ with ‘help.’

The vice-president of a company went to the chairman of the board and said, ‘We are thinking of starting a new program. It will help us increase profits, and it will also help the environment.’

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

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

A question arises as to how people would apply the concepts of intention and acting intentionally to each of these vignettes. Confronted with the harm vignette, would people ordinarily say that it was the chairman's *intention* to harm the environment? Would they say that the chairman *intentionally* harmed the environment? And parallel questions apply to the help vignette. Would people say that the chairman intentionally helped the environment? That it was his intention to help the environment?

The best way to answer these questions would be to run a systematic psychological experiment. That experiment has been completed (Knobe 2004), and the results are as follows:

Percentage saying 'yes'	'Intention'	'Intentionally'
Help	0%	20%
Harm	29%	87%

Looking at this table, one sees two key findings.

First, the results for people's use of the word 'intention.' These are exactly what one would expect. In both conditions, the majority of subjects regarded the outcome as a merely foreseen side-effect and therefore concluded that the agent had no 'intention' to bring it about.

But now turn to the results for 'intentionally.' It appears that people's intuitions as to whether or not the agent performed the behavior *intentionally* do not stand in any simple relation to their intuitions as to whether or not the agent had an *intention* to perform it. In fact, people's use of 'intentionally' does not seem to track any of the agent's mental states. Rather, it appears to be influenced by the *moral* status of the behavior itself. People said that the behavior was performed intentionally when it was morally bad and unintentionally when it was morally good. This basic pattern has been replicated in a number of subsequent experiments (Hauser forthcoming; Knobe & Mendlow forthcoming; McCann forthcoming; Nichols unpublished data).

The asymmetry in people's use of 'intentionally' begins to make sense when we consider people's judgments of praise and blame. There seems to be an asymmetry whereby people are willing to blame the agent for foreseen side-effects that are bad but

not to praise the agent for foreseen side-effects that are good. (So, for example, people blame the chairman for harming the environment but do not praise the chairman for helping the environment; Knobe 2003.) If we suppose that the concept of acting intentionally is best understood as a tool for making judgments of blame and praise, it makes sense that this concept would come to be applied asymmetrically as well.

These are puzzling phenomena, and a number of theories have been proposed to explain them (Adams & Steadman 2004a, 2004b; Knobe forthcoming; Malle & Nelson 2003; Mele 2003; Nadelhoffer forthcoming a, forthcoming b). Here, however, we will not be focusing on the details of these theories. Instead, we will be concerned with the divergence between people's use of 'intentionally' and their use of 'intention.' People appear to be willing to say that the chairman *intentionally* harmed the environment but not that the chairman had an *intention* to harm the environment. This finding shows clearly that people are sometimes willing to say that an agent intentionally performed a behavior even when they are not willing to say that the agent had a corresponding intention.

In this first experiment, each subject received only one question. So some subjects received a question with 'intentionally,' others received a question with 'intention,' and the key findings were obtained by comparing the answers given by these two groups of subjects. It might be thought, however, that subjects would have given different answers if they had each received *both* questions. In particular, it might be thought that subjects would see some kind of contradiction between the statement that the agent acted intentionally and the statement that the agent had no intention to perform the behavior.

One might therefore expect that subjects who received both questions would feel a certain pressure to provide the same answer — either ‘yes’ or ‘no’— in both cases.

To address this worry, McCann (forthcoming) ran a modified version of the experiment in which subjects received both questions on the same questionnaire. The results were exactly the opposite of what one might expect. The gap between subjects’ answers to the two questions was even wider in this new experiment than it had been in the earlier experiments where each subject received only one question. In the new experiment, 80% of subjects said that the chairman harmed the environment *intentionally*; only 12% said that it was his *intention* to harm the environment.

It appears, then, that people are perfectly willing to apply the word ‘intentionally’ in cases where they would not be willing to apply the word ‘intention.’ Our inquiry will be concerned with the question as to what these results show about people’s concept of acting intentionally.

III

We see three plausible answers to this question.

(1) First, it might be said that people’s use of words like ‘intentionally’ does not give us an accurate picture of the concept of acting intentionally – indeed that people sometimes apply the word ‘intentionally’ to behaviors that do not really fulfill the criteria given by their own underlying concept. This view has been put forth by Adams and Steadman (2004a, 2004b), Malle and Nelson (2003) and Nadelhoffer (forthcoming a, forthcoming

b). All of these authors offer sophisticated accounts of the psychological mechanisms that lead people to apply the word ‘intentionally’ (incorrectly, it is claimed) to morally bad side-effects.

We will not be concerned here with questions about whether these theories are right or wrong. Instead, we simply want to emphasize that none of the theories provide any reason to believe that the concept of acting intentionally is more closely connected to the concept of intention than one might otherwise have assumed.

So, for example, Adams and Steadman (2004a, 2004b) have argued the effect described above is due entirely to conversational pragmatics. The idea is that people are unwilling to use the sentence

(a) The chairman unintentionally harmed the environment

because they want to avoid the implicature that the chairman was not to to blame for what he did. This hypothesis is at least a reasonable one (though there is some experimental evidence against it; Knobe 2004). But one still needs an explanation for the fact that people are willing to use the sentence

(b) It was not the chairman’s intention to harm the environment.

Presumably, the difference between these two sentences does not derive entirely from the fact that ‘unintentionally’ is an adverb and ‘intention’ is a noun. There must be some fundamental difference between the concept of acting intentionally and the concept of intention that accounts for people’s willingness to use sentence (b) but not sentence (a).

Similar considerations apply to the work of Malle and Nelson (2003) and Nadelhoffer (forthcoming a, forthcoming b). These authors suggest that people’s

application of the word ‘intentionally’ in the harm vignette is colored by the negative emotions they feel toward the agent. This is certainly a reasonable hypothesis (though, once again, there is experimental evidence against it; Knobe & Mendlow forthcoming; Knobe forthcoming). But one has to ask why people’s emotions have not also affected their application of the word ‘intention.’ The answer must surely be that there is some important difference between ‘intentionally’ and ‘intention.’

Note finally that it is not enough simply to say that, if we wanted to excuse the agent, we would be more likely to use sentence (a) than sentence (b). This may be true, but it only pushes back the question. One wants to know *why* people tend to use the word ‘intentionally’ (rather than ‘intention’) when they are engaged in discussions of praise and blame. The answer, presumably, will have something to do with the difference between the concept of acting intentionally and the concept of intention, and it is that difference that we are trying to understand here.

(2) A second plausible view would be that there is indeed a connection between the concept of acting intentionally and the concept of intention but that this connection is far more complex than one might originally have thought. Proponents of this view do not claim that an action can only be intentional if the agent had an intention to perform it, but they do claim that, for every intentional action, there must be *something* that the agent had an intention to do (Bratman 1984, 1987; Mele 1989, 1992).² So, for example, when the chairman intentionally harms the environment, he does not have an intention to harm

² This seems like the right place to acknowledge our debt to Bratman and Mele. Their seminal work set the stage for all subsequent discussions of these issues, including our own.

the environment, but he does have an intention to implement the program. It might be said, then, that chairman's intention to implement the program is what makes his harming of the environment an intentional action. The hope is that we can articulate a set of principles that will allow us to determine, for each possible intention, which actions that intention would render intentional.

The problem deepens when we consider a broader variety of cases. Thus, suppose that an assassin is trying to shoot the president but believes that it is extremely unlikely that he will succeed. Here it seems a bit odd to say that the assassin has formed an 'intention' to shoot the president, but if he actually did succeed, we would surely say that he had shot the president 'intentionally.' Now, why exactly do we use the term 'intentionally' in cases like these? The obvious way to respond would be to appeal to the fact that the agent was specifically *trying* to perform the behavior that he ended up performing. But now it seems that there is a problem; this sort of response does not in any way involve the concept of intention. The philosopher may therefore be tempted to look around for some kind of intention that can be associated with this behavior. And indeed it is possible to find one — an *intention to try* to shoot the president. So one option would be to say that a behavior can be rendered intentional either by an intention to perform it or by an intention to *try* to perform it. Offering an account along these basic lines, Mele (1992) proposes the following 'protoanalysis':

PA2. S intentionally A-s if and only if S A-s and either (a) S's A-ing is caused in the right way by an intention to A or (b) S performs some action B that is caused in the right way by an intention whose plan component represents S's A-ing as a goal relative to S's intended B-ing and S's B-ing appropriately generates S's A-ing.

At this point, one can get a sense for the role of the concept of intention in analyses of the concept of acting intentionally. Clearly, it is not that philosophers are looking for the simplest, most elegant analysis and then are naturally led to give a prominent role to the concept of intention. On the contrary, the concept of intention seems only to be making the analysis more complex and unwieldy. So the order of justification must be quite different from what it appears to be at first. It seems that philosophers start out with the view that the concept of intention *must* be playing a key role. What makes their analyses so ingenious and impressive is that, despite the massive divergences in use between the concepts *acting intentionally* and *intention*, they are somehow able to analyze one in terms of the other.

In short, the philosophical analyses we have been discussing do not themselves provide strong reason to believe that the concept of acting intentionally is intimately linked to the concept of intention. Their aim is a somewhat different one. What they show is that, if we have some independent reason to think that there is an intimate link between the concept of acting intentionally and the concept of intention, we can hold onto that view even in the face of all of the apparent divergences in use. Presumably, the primary reason for positing such an intimate link would be the morphological relation between the English words ‘intentionally’ and ‘intention.’

(3) A third plausible view would be that the words ‘intention’ and ‘intentionally’ express two entirely different concepts and that there is nothing important to be learned from the fact that they sound so much alike. Proponents of this view would presumably say that moral considerations play no role in the concept of intention but that moral

considerations really do play a role in the concept of acting intentionally (Harman 1976; Knobe 2003, forthcoming; Mele 2003). The basic idea would be that the concept of intention is a purely psychological concept whereas the concept of acting intentionally is, in some fundamental sense, a moral one.

This third view says that the meaning of the word ‘intentionally’ is a kind of primitive. We do not understand the meaning of ‘intentionally’ by understanding the meanings of its component morphemes and then understanding how they fit together to form the meaning of the whole. Rather, we have an independent concept of acting intentionally (distinct from our concept of intention), and we understand the meaning of ‘intentionally’ by understanding that it expresses this concept. The mind includes certain mechanisms for determining whether or not a given behavior was performed intentionally, and it seems likely that these mechanisms make use of various other concepts. But there is a big difference between (a) the relatively banal claim that we use various other concepts to determine whether or not a behavior was performed intentionally and (b) the more controversial claim that the word ‘intentionally’ can actually be *defined* in terms of other concepts. So, for example, suppose that we have an innate ‘moral faculty’ (Dwyer 1999; Harman 1999; Hauser forthcoming; Mikhail et al. 1998) and that this faculty can determine whether or not behaviors were performed intentionally. When we learn the meaning of the word ‘intentionally,’ we might simply be learning to map that word onto a concept that is already being used by the moral faculty. On this model, the moral faculty might be using various other concepts to determine whether or not a behavior was performed intentionally, but the language faculty does not contain a definition of ‘intentionally’ in terms of other concepts.

A key implication of this view is that the relationship between ‘intention’ and ‘intentionally’ is radically different from the relationships between ‘compassion’ and ‘compassionately,’ ‘love’ and ‘lovingly,’ ‘lust’ and ‘lustfully.’ Most linguists think that the meaning of the adverb in each of these pairs is derived from the meaning of the corresponding noun. Our third view claims that the adverb ‘intentionally’ does not work like this. The claim is that the relationship between ‘intention’ and ‘intentionally’ is more like the relationship between ‘ration’ and ‘rationally’ — just two separate words that happen to be morphologically related.

To a first glance, it may appear that this third view is not really plausible at all, but perhaps some of the initial sense of absurdity will dissolve if we look more closely at some of the other adverbs on the list above. It seems that there is a single rule which can be used to derive the meaning of each of these adverbs from the meaning of the corresponding noun. In each case, the adverb is used to indicate that a behavior was performed in the manner of one who has the mental state denoted by the noun. Thus, the word ‘lovingly’ means roughly ‘in the manner of one who feels love,’ and ‘lustfully’ means roughly ‘in the manner of one who feels lust.’ The fact that all of these adverbs seem to follow the same rule gives us some reason to believe that the meanings of the adverbs truly were derived from the meanings of the nouns.

Let us pause for a moment to consider this rule in more detail. It seems clear that none of the adverbs on our list can be used simply to indicate that a behavior was *caused* in the right way by a particular mental state. If a man applies for a job at a restaurant because he is lusting after one of the employees there, it would not normally be correct to say that he fills out the application forms ‘lustfully.’ He could only be correctly said to

act lustfully if he actually behaved in the *manner* of one who feels lust (e.g., by staring at a woman in an especially provocative way). Similar remarks apply to many other adverbs derived from mental state nouns — ‘compassionately,’ ‘lovingly,’ ‘pityingly,’ ‘cheerfully,’ ‘passionately,’ and so on. None of these adverbs serve simply to indicate that a behavior was caused by a given mental state. All of them are used to indicate something about the manner in which the behavior was performed.

But it should be immediately clear that ‘intentionally’ does not follow this rule. A person could knock over a vase intentionally even if she did not knock it over in the manner that would be typical of one who had an intention. Indeed, her behavior could be said to have been performed ‘intentionally’ even if she did everything possible to perform it in the manner of one was not acting on any intention at all. Thus, it seems that ‘intentionally’ is not derived from ‘intention’ by the same rule that is used to derive any of the other adverbs we have discussed.

We come now to a key point about philosophical accounts of the relationship between ‘intentionally’ and ‘intention.’ None of these accounts follow from a *general* theory — a theory that can be used to derive the meanings of a whole class of adverbs from the meanings of a whole class of nouns. All of them are ad hoc accounts that were designed solely to accommodate intuitions about the use of one particular term. For this reason, the view that the meaning of ‘intentionally’ derives from the meaning of ‘intention’ is on much shakier ground than the corresponding view about the various other adverbs we have discussed.

Ultimately, though, this sort of argument is unlikely to be convincing all by itself. The real issue is one of linguistic universality. If the same morphology arose in all other

languages — so that the word for ‘intentionally’ was always derived from the word for ‘intention’ — the best guess would be that there was some fundamental connection between the concept of acting intentionally and the concept of intention. By contrast, if many different languages had words for ‘intentionally’ but almost all of these words were not derived from words for ‘intention,’ we would have reason to suspect that the relationship we find in English is simply a coincidence.

IV

A casual inspection of other languages show a great deal of variation in their words for ‘intentionally’:

- The Slovak word *naschvál* is derived from a verb for ‘to acknowledge’;
- the French *exprès* is related etymologically to a word meaning ‘expression’;
- the Russian *spetsalno* can also be used to mean ‘specifically’;
- the Turkish *bilerek* is derived from a word meaning ‘knowledge’;
- the Vietnamese *co tinh* comes from a word mean ‘to try’ and a word meaning ‘emotion’;
- the Latin *sponte* is derived from a word meaning ‘will’ or ‘volition.’

A question arises, however, as to whether these words really have exactly the same use as the English word ‘intentionally.’ It might turn out, e.g., that the use of the Vietnamese phrase ‘co tinh’ really is connected to the concept of trying in a way that the use of the

English word ‘intentionally’ is not. We will not be examining the use of all of these words and phrases here. Instead, we will focus on a particularly interesting case that arises in Hindi.

The Hindi verb *jaan* means ‘to know.’ One also finds *jaan* at the root of certain nouns, including *jaan-na* (‘knowing’) and *jaan-kari* (‘knowledge’). As far as we can tell, there is no difference in meaning between the Hindi word *jaan* and the English ‘to know,’ though perhaps there is some difference of which we are unaware.

Derived from the noun ‘jaan,’ there is an adverb: *jaan-bujhkar*. We will be concerned with the question as to how this adverb is used. In cases where an agent foresees that he will be bringing about a certain effect but does not actually have an intention to bring about that effect, will Hindi-speakers say that he brought about the effect *jaan-bujhkar*?

To answer this question we conducted a simple experiment. The harm and help vignettes were translated into Hindi by one of the authors. The aim was to determine how Hindi-speakers would apply the words ‘jaan’ and ‘jaan-bujhkar’ to each of these vignettes.

Subjects were 61 Hindi-speaking students in South Asian clubs at Princeton University and Yale University. Each subject was randomly assigned either to the *harm condition* or the *help condition*. Subjects in the harm condition were given the harm vignette; subjects in the help condition were given the help vignette. Within each of these conditions, subjects were randomly assigned to either the *jaan-bujhkar condition* or the *jaan condition*. Subjects in the *jaan-bujhkar condition* were asked (in Hindi): ‘Did the chairman harm [help] the environment *jaan-bujhkar*?’ Subjects in the *jaan condition* were

asked: ‘Did the chairman jaan that starting the program would harm [help] the environment?’

Crossing the variable of vignette type (harm vs. help) with the variable of question type (jaan-bujhkar vs. jaan), we obtain four basic conditions. The percentage of subjects answering ‘yes’ in each of these conditions is given below:

	<i>jaan</i>	<i>jaan-bujhkar</i>
help	90%	14%
harm	80%	75%

In this pattern of results, we again see two key findings.³

First, looking at the column on the right, we see that use of the Hindi ‘jaan-bujhkar’ closely mirrored use of the English word ‘intentionally,’ with people saying that the chairman harmed the environment jaan-bujhkar but not that he helped the environment jaan-bujhkar. This finding suggests that the Hindi word ‘jaan-bujhkar’ expresses the same concept as the English word ‘intentionally.’⁴

Second, looking at the top row, one sees that use of the word ‘jaan-bujhkar’ diverged from use of the word ‘jaan.’ People said that the chairman did jaan that he would be helping the environment but that he did not help the environment jaan-bujhkar.

³ Detailed statistical analyses are reported in the appendix.

⁴ This finding provides some preliminary support for the conclusion that the pattern described here is a linguistic universal. If the pattern does in fact turn out to be universal, we will have at least prima facie evidence for the thesis that it is the product of an innate, domain-specific mechanism. (Additional support for this thesis comes from recent studies showing that the pattern emerges even in the responses of four-year-old children; Leslie, Knobe & Cohen forthcoming.)

This second finding indicates that the word ‘jaan-bujhkar’ is not simply an adverbial form of ‘jaan.’

Overall, then, the results point to a surprising conclusion about the relationship between the words ‘intentionally’ and ‘jaan-bujhkar.’ It appears that, despite their radically different morphologies, these two words are expressing the same basic concept. In English, this concept is expressed by a word derived from ‘intention.’ In Hindi, it is expressed by a word derived from ‘jaan.’ But the use of the concept does not closely mirror use of either of these words. Rather, use of the concept depends in part on the *moral* status of the behavior — so that its use is closer to that of ‘intention’ for morally good behaviors and closer to ‘jaan’ for morally bad behaviors.

V

Our inquiry has been concerned with the relationship between the concept of acting intentionally and the concept of intention. It is normally assumed that these two concepts are related in some fundamental way. We wanted to see whether there was any real evidence for that assumption — evidence that the concept of acting intentionally is more closely related to the concept of intention than it is to the concepts of wanting, trying, foreseeing, and so forth.

When we looked at the ways in which the two concepts were actually used, we found little direct evidence of a connection. People’s use of the concept of acting intentionally seemed to depend on the moral goodness or badness of the behavior concerned. As long as the behavior was morally good, use of the concept of acting

intentionally was closely connected to use of the concept of intention. But when the behavior was morally bad, use of the concept of acting intentionally more closely resembled use of the concept of foresight. This result seemed to point to a key difference between the concept of acting intentionally and the concept of intention.

But although the use of the two concepts does not strongly suggest that they are linked, we may appear to have another source of evidence: the English word ‘intentionally’ is obviously and conspicuously derived from the English word ‘intention,’ and one might therefore feel that the concepts these words express must be related as well.

There is certainly something intuitively appealing about this argument. The only problem is that it relies on certain contingent features of the English language. Turning to Hindi, we find that the word for ‘intentionally’ is derived from the word for ‘know.’ Thus, if analytic philosophy were practiced primarily by Hindi-speakers, the prevailing view would almost certainly have been that the concept of acting intentionally was connected in some fundamental way with the concept of knowledge. Researchers would then have faced the opposite sort of puzzlement from the one facing English-speaking researchers today. They might well have wondered how it could be that, for morally good behaviors people sometimes said that an agent did know (jaan) that he was performing the behavior but nonetheless did not perform the behavior intentionally (jaan-bujhkar).

Looking at these hypothetical philosophers from our present perspective, we can be almost certain that they are making a mistake. We see clearly that the Hindi language just happens to show a morphological connection between the word for ‘intentionally’ and the word for ‘know’ but that there are other languages in which the words for these

concepts are not morphologically connected and that, ultimately, there is no reason to suppose that the concept of acting intentionally is any more closely related to the concept of knowledge than it is to certain other mental state concepts.

A question now arises as to whether a similar argument should be applied to the morphology of English. It is certainly true that the word ‘intentionally’ is morphologically connected to the word ‘intention.’ But in light of the evidence from people’s use of the concept in English and other languages, should we simply conclude that the morphology of the word ‘intentionally’ is misleading us about the concept it expresses? Perhaps future research will show that we have good reason to resist this conclusion, but at this point, we see no convincing evidence for the view that the concept of acting intentionally has any special connection to the concept of intention.

Appendix

In this appendix, we provide more detailed methodological and statistical information about the experiment reported in Section III. The information provided here may prove helpful to some readers but is not essential to an understanding of the chief conceptual and philosophical points of the paper.

Subjects were 61 people attending events organized by South Asian student groups at Princeton University and Yale University. Each subject was approached by the experimenter and asked whether he or she knew Hindi. Those who answered ‘yes’ to this question were given a brief questionnaire. After filling out the questionnaire, subjects were thanked and debriefed.

The experiment used a 2 x 2 design. One factor was *vignette-type* (harm vs. help); the other was *question-type* (jaan vs. jaan-bujhkar). Both factors were between-subjects. Twice as many subjects were assigned to the jaan-bujhkar conditions as to the jaan conditions; equal numbers of subjects were assigned to the harm and help conditions.

Each subject received a questionnaire containing a short vignette followed by two questions. The questionnaires varied across conditions as follows:

harm/jaan: [harm vignette]

‘Did the chairman jaan that implementing the policy would harm the environment?’

‘Does the chairman deserve blame for harming the environment?’

harm/jaan-bujhkar: [harm vignette]

‘Did the chairman harm the environment jaan-bujhkar?’

‘Does the chairman deserve blame for harming the environment?’

help/jaan: [help vignette]

‘Did the chairman jaan that implementing the policy would help the environment?’

‘Does the chairman deserve praise for helping the environment?’

help/jaan-bujhkar: [help vignette]

‘Did the chairman help the environment jaan-bujhkar?’

‘Does the chairman deserve praise for helping the environment?’

All questions were translated into Hindi. The order of questions was counterbalanced.

The following table displays the percentages of subjects answering ‘yes’ to the questions about jaan-bujhkar and jaan:

	<i>jaan</i>	<i>jaan-bujhkar</i>
help	90%	14%
harm	80%	75%

Subjects were significantly more likely to say that the chairman acted jaan-bujhkar in the harm condition than in the help condition, $\chi^2(1, N=41) = 15.3, p < .001$. Similarly, subjects were significantly more likely to say that the chairman did jaan that the environment would be helped than that he helped the environment jaan-bujhkar, $\chi^2(1, N=31) = 16.3, p < .001$. These comparisons remain significant even after one applies the Bonferroni adjustment, yielding an alpha level of .025.

The following table reports the percentages of subjects in each condition who said that the chairman deserved praise or blame.

	<i>jaan-bujhkar</i>	<i>jaan</i>
help	24%	30%
harm	90%	80%

To analyze these data we constructed a logit model ($N=61$) with praise or blame as a dependent variable and vignette-type (harm vs. help) and question-type (jaan vs. jaan-bujhkar) as independent variables. There was a significant main effect for vignette-type, $Z=4.0$. However, there was no significant effect of question-type, $Z=0.3$, nor was there a significant interaction between the two variables, $Z=0.9$.

Judgments about whether the chairman acted jaan-bujhkar were correlated with judgments about whether the chairman deserved praise or blame, $r=.49$, $p=.001$.

Judgments about whether the chairman did jaan that starting the program would help or harm the environment were not significantly correlated with judgments of praise or blame, $r=-.38$, $p=.10$.

Bibliography

- Adams, F. & Steadman, A. (2004a). Intentional action in ordinary language: Core concept or pragmatic understanding? *Analysis*, 64, 173-181.
- Adams, F. & Steadman, A. (2004b). Intentional action and moral considerations: Still pragmatic. *Analysis*, 64, 268-276.
- Austin, J. L. (1979). Three ways of spilling ink. In *Philosophical Papers* (3rd ed.) Oxford: Oxford University Press.
- Bratman, M. (1984). Two faces of intention. *Philosophical Review*, 93, 375-405.
- Bratman, M. (1985). Davidson's theory of intention. In B. Vermazen and M.B. Hintikka (Eds.). *Essays on Davidson. Actions and Events*. Oxford: Clarendon Press.
- Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.
- Davidson, D. (1973). Freedom to act. In *Essays on Freedom of Action*, ed. Ted Honderich. London: Routledge and Kegan Paul.
- Dwyer, S. (1999). Moral competence. In K. Murasugi and R. Stainton (Eds.), *Philosophy and Linguistics*. Westview Press.
- Harman, G. (1976). Practical reasoning. *Review of Metaphysics*, 29, 431-463.
- Harman, G. (1999). Moral philosophy and linguistics. In K. Brinkmann (Ed.), *Proceedings of the 20th World Congress of Philosophy: Volume 1: Ethics*. Philosophy Documentation Center, 107-115.
- Hauser, M. (forthcoming). *Moral minds: The unconscious voice of right and wrong*. NY: Harper Collins.

- Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63, 190-193.
- Knobe, J. (2004). Intention, intentional action and moral considerations. *Analysis*, 64, 181-187.
- Knobe, J. (forthcoming). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies*.
- Leslie, A., Knobe, J. & Cohen, A. (forthcoming). Acting intentionally and the side-effect effect: 'Theory of mind' and moral judgment. *Psychological Science*.
- Malle, B. F., & Nelson, S. E. (2003). Judging mens rea: The tension between folk concepts and legal concepts of intentionality. *Behavioral Sciences and the Law*, 21, 563-580.
- McCann, H. (forthcoming). Intentional action and intending: Recent empirical studies. *Philosophical Psychology*.
- Mele, A. (1989). Intention, belief, and intentional action. *American Philosophical Quarterly*, 26, 19-30.
- Mele, A. (1992). *Springs of Action*. Oxford: Oxford University Press.
- Mele, A. (2003). Intentional action: controversies, data, and core hypotheses. *Philosophical Psychology*, 16, 325-340.
- Mikhail, J., Sorrentino, C., & Spelke, E. S. (1998). *Toward a universal moral grammar*. Paper presented at the Proceedings of the Cognitive Science Society.
- Nadelhoffer, T. (forthcoming a). Praise, side effects, and intentional action. *The Journal of Theoretical and Philosophical Psychology*.

Nadelhoffer, T. (forthcoming b). Blame, badness, and intentional action: A reply to
Knobe and Mendlow. *The Journal of Theoretical and Philosophical Psychology*.