Forthcoming In Hutto, D. & Ratcliffe, M. (ed.)
Folk psychology reassessed. Kluwer/Springer Press.

Folk Psychology: Science and Morals[1]

Joshua Knobe

It is widely agreed that folk psychology plays an important role in people's moral judgments. For a simple example, take the process by which we determine whether or not an agent is morally blameworthy. Although the judgment here is ultimately a moral one, it seems that one needs to use a fair amount of folk psychology along the way. Thus, one might determine that an agent broke the vase *intentionally* and therefore conclude that she is *blameworthy* for breaking it. Here it seems that one starts out with a folk-psychological judgment (that the agent acted intentionally) and then uses it as input to a process that eventually yields a moral judgment (that the agent is blameworthy). Many other cases have a similar structure.

In recent years, however, a number of studies have shown that there are also cases in which the arrow of causation goes in the opposite direction. That is, there appear to be cases in which people start out with a *moral* judgment and then use it as input to a process that eventually yields a *folk-psychological* judgment (Knobe 2003a, 2003b, 2004, 2005a, 2005b). These findings come as something of a surprise, and it can be difficult to know just what to make of them.

My own view is that the findings are best explained by the hypothesis that moral considerations truly do play a role in people's underlying folk-psychological concepts (Knobe 2003b, 2004, forthcoming). The key claim here is that the effects revealed in recent experiments are not the result of any kind of 'bias' or 'distortion.' Rather, moral considerations truly do figure in a fundamental way in the issues people are trying to resolve when they grapple with folk-psychological questions.

I must confess, however, that not all researchers in the field share this view. Although many have been convinced that moral considerations actually do play a role in folk-psychological concepts, others have suggested that there might be better ways to account for the results of recent experiments. What we are left with, then, is an

[1] I am grateful to the editors for extensive comments on an earlier draft.

increasingly complex debate. Critics of my original proposal have constructed alternative hypotheses that seem to account for all of the data without assigning any fundamental role to moral considerations. Defenders then conduct new experiments that appear to falsify these alternative hypotheses. But the critics inevitably respond by constructing even more sophisticated alternative hypotheses that manage to explain all of the new data while still assigning no fundamental role to moral considerations. And so the debate continues, with each new iteration yielding new theoretical insights and empirical discoveries.[2]

I will not be continuing that debate here. Instead, I want to focus on an issue that is somewhat broader and perhaps more basic. The critics sometimes seem to feel that moral considerations just *couldn't* be playing a fundamental role in folk psychology. The feeling is that, independent of the merits of any particular alternative explanation, one can tell that there must be *some* way to construct a valid alternative. This feeling is never articulated explicitly. Still, it comes through in the palpable sentiment that my defenders and I are upholding an absurd view and that we had really better come back to our senses.

My aim here is to confront that sentiment head on. In the first section, I briefly review experimental evidence that suggests that people's moral judgments can sometimes affect their folk-psychological judgments. Then, in the second section, I ask whether we have any general theoretical reasons to expect that moral considerations will not play any fundamental role in folk-psychological concepts.

I

Let us turn, then, to three folk-psychological concepts whose application has been studied experimentally. The first two have already been discussed in earlier papers and will only be described here in a highly condensed summary form. The third appears here for the first time, and I therefore discuss it in greater detail.

[2] For some contributions to this debate, see Adams and Steadman (2004a, 2004b, forthcoming), Harman (forthcoming), Malle (forthcoming), McCann (forthcoming), Meeks (forthcoming), Mele (2003), Morton (forthcoming), Nadelhoffer (2004, 2005), Nichols and Ulatowski (2006), Sverdlik (forthcoming), Turner (forthcoming), Yoo (forthcoming), Young et al. (forthcoming).

Warning: These next two subsections simply summarize my earlier work on intentional action and reason explanations. Readers who are already familiar with that work should skip directly to the subsection on 'Valuing.'

Intentional action

People ordinarily distinguish between behaviors that are performed *intentionally* (e.g., hammering in a nail) and those that are performed *unintentionally* (e.g., accidentally bringing the hammer down on one's own thumb). Clearly, this distinction sometimes has important implications for questions about moral praise and blame, but it is usually assumed that the distinction itself is a purely psychological one. Nonetheless, an ever-growing body of experimental evidence indicates that the moral status of a behavior can actually have an impact on whether or not people regard it as intentional.

The best way to demonstrate this influence of moral judgments on ascriptions of intentional action is to construct pairs of cases that are almost exactly alike but that differ in their moral status. Here is the first element in one such pair:

> 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.

Faced with this first case, most people say that the chairman *intentionally* harmed the environment.

But now suppose that we create a morally good version by simply 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.

When given this second case, most people say that the chairman *unintentionally* helped the environment. Yet the two cases are identical in almost all respects. It seems that the only major difference between them lies in the moral status of the agent's behavior.

In the years since this result was first reported, it has been replicated and extended in a wide variety of additional experiments. It has been shown that the effect continues to emerge when the stories are translated into Hindi and run on Hindi-speaking subjects (Knobe & Burra forthcoming), when the stories are simplified and given to subjects who are only four years old (Leslie et al. 2005), and even when the stories are given to subjects who have deficits in emotional processing due to frontal lobe damage (Young, et al. forthcoming).[3] At this point, no one doubts that people's use of the word 'intentionally' really is influenced by their moral judgments. The debate is simply about what this effect can tell us about the nature of folk psychology.

Reason explanations

Faced with this evidence that moral considerations play a role in people's application of the concept of intentional action, one possible response would be to deny that the concept of intentional action truly is a part of folk psychology. This response would allow us to hold on to the idea that morality plays no role in folk psychology, albeit at the expense of forcing us to admit that our intuitive notion of the scope of folk psychology was not quite correct. To me at least, this response seems a bit desperate, and no one has actually argued for it in print. Still, it comes up often in conversation, and as experimental research continues to show new ways in which the concept of intentional action is sensitive to moral considerations, it may come to seem more and more plausible.

It can be shown, however, that similar effects arise even for concepts that are undeniably folk-psychological. Thus, consider the practice of explaining behavior using *reasons*. A clear example would be the sentence:

[3] For the original experiment, see Knobe 2003a. For further replications and extensions, see Adams and Steadman (forthcoming), Knobe (2003b, 2004), Knobe and Mendlow (forthcoming), Nadelhoffer (2005, forthcoming a, forthcoming b), Malle (forthcoming), McCann (forthcoming), and Nichols and Ulatowski (2005).

He went to the kitchen to get a beer.

This sentence explains an agent's behavior ('He went to the kitchen…') by giving his reason for performing it ('…to get a beer'). Here we seem to have a prototypical case of a folk-psychological judgment. No one would claim that explanations of this type belong to the domain of moral cognition.

And yet, it can be shown that moral judgments actually affect people's use of reason explanations (Knobe 2004). Indeed, the effect can be seen in the very same pair of vignettes we used above. Faced with the first vignette, most people think it sounds right to say:

The chairman harmed the environment in order to increase profits.

But faced with the second vignette, most people *don't* think it sounds right to say:

The chairman helped the environment in order to increase profits.

This pattern of results suggests that people's use of reason explanations is actually sensitive to moral considerations.

It is not known precisely why this effect arises. One plausible hypothesis would be that people are using the concept of intentional action in the process by means of which they evaluate reason explanations. Perhaps people only accept reason explanations for behaviors that they regard as intentional. Then, since moral considerations play a role in people's concept of intentional action, they end up playing a role (somewhat indirectly) in the practice of reason explanation.

Valuing

People ordinarily distinguish between *desiring* and *valuing*. Thus, when a heroin addict is roaming the streets looking for his next fix, we might say that he 'wants' the fix but not that he 'values' it. And we would say the same about the man on a diet who feels overwhelmed by an urge to have another slice of chocolate cake. Philosophers typically find that they all share the same intuitions about how to apply the concept of valuing in cases like these, but it has proved notoriously difficult to say anything very definite about the basic criteria underlying these intuitions. One wants to know exactly how people go about distinguishing values from attitudes of other kinds.

This question has not received much attention from researchers in folk psychology, but it has been discussed extensively within a certain tradition in moral philosophy. This tradition begins with Watson's influential claim that

> an agent's values consist in those principles and ends which he — in a cool and non-self-deceptive moment — articulates as definitive of the good, fulfilling and defensible life. (Watson 1975: 215)

Watson later retracted that claim, worrying that it conflated the notion of valuing with the notion of judging something to be good (Watson 1987). But in the years that followed, a number of other philosophers have offered competing accounts.[4] We will not be concerned here with the differences among these various proposals. Instead, the focus will be on the assumption, shared by all of the views proposed thus far, that the concept of valuing can be defined in purely descriptive, non-normative terms.

I had never thought to question this assumption until the philosopher Erica Roedder suggested to me (in conversation) that there might be more to the story. She pointed out that the ordinary distinctions between desiring and valuing might be bound up in a fundamental way with certain *moral* questions. So, for example, when we are trying to determine whether or not the heroin user 'values' his next fix, it might be that we are not simply concerned with purely descriptive questions about the nature and functional role of the user's attitude. Perhaps our reluctance to classify this attitude as a 'value' is due in part to our sense that heroin truly *is* a bad thing.

One way to make sense of this hypothesis is to suppose that the concept of valuing is a prototype concept. In other words, we can suppose that the concept of valuing is represented by a cluster of features, such that no individual feature is strictly necessary but each feature has been assigned a certain weight. If a particular attitude shows enough of the relevant features, it will be classified as one of the agent's 'values.' It would be extremely difficult to provide an exhaustive list of the features that play a role here, but we can easily list a few that are likely to be relevant. When people are trying to

[4] See especially Bratman (2000), Copp (1995), Lewis (1989) Sayre-McCord and Smith (2005) and Smith (1994). As far as we know, the only previous empirical studies of people's use of the concept are the excellent experiments in Malle and Edmondson (2006).

determine whether or not the agent values a certain object *o*, they probably consider psychological features like:

- whether the agent has a conscious belief that *o* is good
- whether the agent is motivated to promote *o*
- whether the agent experiences guilt when she fails to promote *o* in circumstances where she could have
- whether the agent has a second-order desire for *o* (i.e., a desire to desire *o*)

Each of these psychological features has a certain weight. But the psychological features are not the only features of the concept. There is also a moral feature, namely, *whether the object o truly is morally good*.

Now, clearly, it would be foolish to suggest that moral goodness is a necessary condition in our concept of valuing. But that is not the claim under discussion here. The claim is simply that moral goodness has a certain *weight* in the process of classification. If an agent has all of the relevant psychological features, this extra weight simply won't be needed. The psychological features prove sufficient all by themselves. So the only way to see the significance of the moral feature is to look at cases where the agent has some of the psychological features but lacks others. In cases like these, the psychological features will not be sufficient all by themselves. The attitude needs the moral feature before it has enough weight to push our intuitions over the critical threshold.

Together, Roedder and I conducted an experiment to test this hypothesis. All subjects were given a story about an agent who has some of the relevant psychological features but lacks others. (In our story, the agent has motivation and guilt but not conscious belief or second-order desire.) The key question was whether people's classification of the agent's attitude would be influenced in any way by the perceived moral status of its object.

Subjects in one condition were given a story in which the agent feels a certain pull toward actions that would normally be perceived as *morally good*:

> George lives in a culture in which most people are extremely racist. He thinks that the basic viewpoint of people in this culture is more or less correct. That is, he believes that he ought to be advancing the interests of people of his own race at the expense of people of other races.

> Nonetheless, George sometimes feels a certain pull in the opposite direction. He often finds himself feeling guilty when he harms people of other races. And sometimes he ends up acting on these feelings and doing things that end up fostering racial equality.
>
> George wishes he could change this aspect of himself. He wishes that he could stop feeling the pull of racial equality and just act to advance the interests of his own race.

After reading this story, subjects were asked whether or not they agreed with the sentence: ‘Despite his conscious beliefs, George actually values racial equality.’

Subjects in the other condition were given a story that was very similar to the first one but in which the agent feels a pull towards actions that would normally be perceived as *morally bad*:

> George lives in a culture in which most people believe in racial equality. He thinks that the basic viewpoint of people in this culture is more or less correct. That is, he believes that he ought to be advancing the interests of all people equally, regardless of their race.
>
> Nonetheless, George sometimes feels a certain pull in the opposite direction. He often finds himself feeling guilty when he helps people of other races at the expense of his own. And sometimes he ends up acting on these feelings and doing things that end up fostering racial discrimination.
>
> George wishes he could change this aspect of himself. He wishes that he could stop feeling the pull of racial discrimination and just act to advance the interests of all people equally, regardless of their race.

These subjects were then asked whether or not they agreed with the sentence: ‘Despite his conscious beliefs, George actually values racial discrimination.’

This experiment provides an initial test of our hypothesis. The attitudes depicted in the two stories differ in their moral significance, but they seem not to differ in any of the relevant psychological features. In both cases, the agent has motivation and guilt but not conscious belief or second-order desire. Yet, despite this similarity in psychological features, we find a marked asymmetry in people’s intuitions. Subjects were significantly more inclined to say that the attitude was one of the agent’s values in the morally good case than they were in the morally bad case. This result provides some tentative support for the view that moral judgments actually do play a role in people’s concept of valuing.

Summing Up

The results described here appear to indicate that people's applications of folk-psychological concepts can sometimes be influenced by their moral judgments. These results therefore provide some initial support for the claim that moral judgments are actually playing a role in people's folk-psychological concepts themselves.

But, of course, one cannot infer directly from the conditions under which a concept is applied to the structure of the concept itself. It is always possible that we will be able to come up with an alternative explanation that accommodates all of the relevant data without according any fundamental role to moral considerations in our underlying folk-psychological concepts. Perhaps the results described above are simply due to conversational pragmatics, emotional biases, or some other factor that has nothing to do with the underlying structure of people's concepts. A number of researchers are actively pursuing explanations along precisely these lines (see, e.g., Adams forthcoming; Malle forthcoming; Nadelhoffer 2004, forthcoming a; Nichols & Ulatowski 2006), and nothing I have said here provides any evidence against their hypotheses. Ultimately, the only way to assess these alternative explanations is to engage in a detailed examination of the existing experimental data.[5]

I will not be taking up that task here. Instead, I will be concerned with the initial motivation that leads researchers to search for alternative explanations in the first place. There seems to be a widespread intuition that moral considerations just *couldn't* be playing any fundamental role in people's folk-psychological concepts and that it therefore *must* be possible to find some other way of explaining the data. This intuition does not appear to depend on the evidence for any particular alternative hypothesis. It seems to stem instead from a more general theoretical commitment.

Clearly, the commitment here is not to the idea that moral considerations never play a fundamental role in any of our concepts. It is usually assumed that moral

[5] For evidence against particular alternative explanations, see Knobe (2004, forthcoming a), Knobe and Mendlow (2005), and Young et al. (forthcoming). Advocates of specific alternative explanations have also provided interesting and important evidence against competing alternative explanations (see especially Nadelhoffer forthcoming b; Nichols & Ulatowski 2006). Note that all of this evidence is directed against explanations that were offered a number of years ago. More recent research has led to the construction of new alternative explanations that accommodate all of the existing data while still assigning no fundamental role to moral considerations in people's folk-psychological concepts (e.g., Malle forthcoming; Nichols & Ulatowski 2006). It will be interesting to see how these new explanations fare in accounting for the results of future experimental studies.

considerations do play a role in the concepts of blameworthiness, fairness, etc., and the researchers pursuing alternative explanations for the data described here do not seem to feel compelled to search for alternative explanations in those other cases as well. So the thought seems to be that there is something special about folk-psychological concepts in particular which makes it implausible that moral considerations could play any fundamental role in them. What I want to ask now is whether there really are any general theoretical reasons for holding this view.

II

Much of the attractiveness of the view appears to stem from the idea that folk psychology is in some important way similar to *science*. This idea is never spelled out explicitly, but the underlying argument seems to run something like this:

(1) Folk psychology is similar in many ways to a scientific theory.

(2) Scientific theories do not classify objects based on their moral properties.

We therefore have good reason to suppose that:

(3) Folk psychology does not classify objects based on their moral properties.

Of course, this is a not deductively valid argument, but it is a powerful one all the same. Both of the premises seem initially plausible, and together they appear to provide strong evidence for the conclusion.

To get a sense for the basic idea behind premise (2), it may be helpful to consider an example. Suppose we were able to observe a team of physicists studying the trajectories of certain projectiles. We might expect them to classify a projectile in terms of its mass, velocity, direction, and so forth. But suppose we then discover that their judgments can actually be influenced in some subtle way by *moral* properties, so that they sometimes end up applying scientific concepts to a projectile differently depending on whether they believe that it was morally right or morally wrong to launch it in the first place. In such a case, we surely would not conclude that moral properties actually play some important role in the basic concepts of physics. Instead, we would assume that the physicists were subject to some kind of bias that distorted their scientific judgment.

In thinking about cases like these, we brush up against some difficult questions about the relationship between science and morals. Someone might argue that initial impressions are deceiving here and that there really is some subtle sense in which scientific theories end up classifying objects on the basis of their moral properties. Perhaps there actually is something to this charge, but let us put it to the side for the moment. For the sake of argument, we can simply assume that scientific theories do not classify objects on the basis of their moral properties. Then we can go on to ask what implications this putative fact about scientific theories might have for the study of folk psychology.

The key move, then, is from the claim that moral considerations are excluded from certain aspects of scientific theorizing to the claim that moral considerations are excluded from parallel aspects of folk psychology. This move rests on a certain analogy between science and folk psychology. The view is that, although science is more rigorous, more systematic and more explicit, we have reason to expect that the most basic practices associated with science will be found in folk psychology as well.[6]

It is this view that I want to examine here. To address these issues, we need to look more closely at the role science plays in people's lives and the factors that have made it such a dominant approach to systematic inquiry. Then we can check to see whether those same factors can be found in the case of folk psychology or whether folk psychology differs from science in some important respect.

1. Contemporary enthusiasm for the analogy between folk psychology and science appears to stem, at least in part, from the extremely salient position that science occupies in modern life. Everywhere one looks, one finds the fruits of scientific inquiry, and it is easy to find oneself thinking that the practices we now associate with science are in some

[6] The idea that folk psychology resembles a scientific theory was perhaps first developed by Churchland (1981) and then came to play an important role in 'theory of mind' research as a result of work by Gopnik and Wellman (1992), Gopnik and Meltzoff (1997), and others. Since that time, numerous researchers have argued that folk psychology does not use quite the same *methods* that we find in scientific inquiry, but almost all of these researchers have assumed that folk psychology should be understood in terms of the typical *goals* of science (prediction, explanation, etc.). In a number of recent works, however, this assumption has been called into question. See especially Andrews (2006), Hutto (2004), Morton (2003) and Wilkes (1981).

way ‘natural’ to human beings. One almost finds it difficult to imagine any other way of generating predictions or explanations.

But, of course, the matter is not so simple. Many of the practices that we now associate with science arose in a particular cultural context in the not-too-distant past. These practices are now quite widespread, but one cannot therefore infer that they reflect anything fundamental about human nature. It may well be that they only came to occupy such a salient position in our society because they do such a good job of solving the kinds of problems we most often encounter in modern life.

Perhaps some of the confusion here arises from our tendency to lump together a diverse array of practices and label them all collectively as ‘science.’ Some of the practices that fall under this label really do seem to reflect something fundamental about human nature. These practices can be found in young children and in people from other cultures, and many cognitive scientists believe that they have an innate basis (see, e.g., Bloom 2004; Gopnik et al. forthcoming; Keil 1989; Pinker 1997). But not all of the practices associated with science work like that. Some of them were only developed in recent centuries and appear to be passed down from one generation to the next through explicit instruction. There is little reason to suppose that these practices reflect anything fundamental about our innate cognitive endowments (Faucher et al. 2002; McCauley 2000).

The thing to keep in mind in discussing practices of this latter type is that they arose as a result of certain contingent historical events. There is an important sense in which the ‘scientific revolution’ of the 16th and 17th centuries truly was a *revolution*. It introduced genuinely new practices, practices that cannot be found in earlier eras. These practices subsequently assumed a dominant role in the kinds of inquiry conducted in systematic research programs, and we have ample evidence that they do a wonderful job of helping us get at the truth about certain difficult questions. But it would be wrong to suppose that there is something basic about human nature that compels us to adopt these practices in the form in which they presently exist. At other times and in other cultures, people have generated predictions using approaches that differed in various ways from the approach we now associate with science.

With this background in place, we can return to our central question. That question was whether we have any general theoretical reason to suppose that folk psychology treats moral considerations in the same way that science does.

2. The idea that folk psychology might be similar to science has been encouraged by the claim that folk psychology should be understood (in a certain technical sense) as a *theory*. The association here is understandable. As soon as one hears the word ‘theory,’ one immediately thinks of the sciences. So when one is told that folk psychology itself should be understood as a theory, one naturally leaps to the conclusion that folk psychology should be understood as something like science. It is therefore essential to remember that the word ‘theory’ was first introduced into this discussion in a highly specialized sense that did not carry any implications about all of the practices we normally associate with science.

The idea that folk psychology should be understood as a theory was first developed by Sellars (1956) and then entered the world of cognitive science through the influential work of Premack and Woodruff (1978). These researchers were concerned with the fact that folk psychology doesn’t just give us a collection of empirical generalizations about observable phenomena but actually provides a deeper sort of account that works by explaining observable behaviors in terms of unobservable mental states. As Premack and Woodruff put it:

> In saying that an individual has a theory of mind, we mean that the individual imputes mental states to himself and to others… A system of inferences of this kind is properly viewed as a theory, first, because such states are not directly observable, and second, because the system can be used to make predictions, specifically about the behavior of other organisms. (Premack & Woodruff 1978:515)

I have no objections to this use of the term ‘theory,’ but when the term is used in this way, one cannot simply assume that every theory is best understood on the model of science. After all, a system of belief can easily qualify as a ‘theory’ in Premack and Woodruff’s sense even if it does not have many of the properties we normally associate with scientific inquiry. To take a particularly glaring example, certain *religions* posit unobservable entities that can be used to predict observable events and might therefore be described as ‘theories.’ Now, it does seem fair to say that a religion can offer us a theory about how the world works, but one sees immediately that the theories offered by religions differ from scientific theories in a number of important respects.

In particular, the argument sketched above seems to depend in a crucial way on the distinctive features of *scientific* theories. There is some intuitive plausibility to the inference: ‘Folk psychology is similar to science. Therefore, it does not classify objects based on their moral properties.’ But the argument loses all its force when we change it to: ‘Folk psychology is similar to religion. Therefore, it does not classify objects based on their moral properties.’ Religions serve a great many different functions in our lives, and prediction is just one. No one would be surprised to find that religious theories are connected in an essential way with moral considerations.

In short, it is easy to get confused by the claim that folk psychology is a ‘scientific theory.’ We really need to divide this claim into two parts — the claim that folk psychology is a *theory* and the claim that folk psychology is *scientific*. The claim that folk psychology is a theory simply isn’t very relevant to the questions we are trying to address here. What we really want to know is whether folk psychology is, in the relevant sense, scientific.

3. Our concern, then, is with the distinctive features of scientific theories — the features that distinguish scientific theories from theories of other types. It seems that these features lie not so much at the level of content as at the level of methodology. The methods we use to evaluate scientific theories seem to differ in some important respects from the methods we use to evaluate theories of other types.

Perhaps the most striking aspect of scientific methodology is its sensitivity to empirical evidence. We use scientific theories to generate predictions which can then be tested through observation or experiment. Theories that yield false predictions may be revised or abandoned. So one way to determine whether folk psychology is something like a scientific theory would be to ask whether it, too, is sensitive in the right way to empirical evidence.

A whole industry of research has arisen to answer this question, and a wide variety of competing theoretical frameworks have now been proposed. Some have argued that people can revise the basic framework of folk psychology using the very same psychological processes that scientists use to revise their theories (e.g., Gopnik & Wellman 1992); others argue that the basic framework underlying folk psychology is innate and is only sensitive to empirical considerations through a process of evolution by natural selection (e.g., Baron-Cohen 1995); and still others have suggested that folk psychology might be subserved by an innate module that uses empirical evidence to set certain highly specific parameters (Stich & Nichols 1998). The debate among these various positions is still ongoing.

But I worry that this research does not really get at the question we are trying to address here. It is not as though scientific theories are the only systems of thought that prove sensitive to empirical considerations. One finds at least some level of sensitivity to empirical considerations even in systems of thought that are clearly non-scientific. Consider a simple example. In the seventeenth century, many European Jews believed that Shabbatai Zvi was the messiah. They then received a shocking piece of disconfirming empirical evidence (Shabbatai Zvi converted to Islam), and most of them soon abandoned their previous belief. What we have here is a clear case of a group of people revising their views in light of empirical evidence. But no one would suggest that the followers of Shabbatai Zvi were propounding a genuine scientific hypothesis! Clearly, their belief was a religious doctrine, and the criteria used to evaluate it therefore differed quite radically from the criteria we typically find in scientific inquiries.

The key mistake here is to assume that we can figure out what is special about scientific inquiry simply by looking at the considerations that scientists normally take into account. This approach has undoubtedly yielded many important insights, but it is not sufficient all by itself. We also need to look for kinds of considerations that scientists *don't* take into account. That is, we need to look for kinds of considerations that figure prominently in other systems of thought but do not play any role in scientific inquiries.

To get a sense for what I mean here, consider the many kinds of criteria we might use in deciding between competing religious doctrines. It seems that many of these criteria play no role at all in scientific investigations. Indeed, one of the key turning points in the scientific revolution was the struggle to establish a special realm of inquiry from which these other criteria would be completely excluded.

For present purposes, one of the most important distinctions between scientific and non-scientific theories lies in the differing roles they assign to *moral* considerations. We expect a religious doctrine to give us some measure of moral guidance, and if it fails to do so, we regard it as deficient in an important respect. By contrast, when we are evaluating a scientific theory, it seems that we are *not* supposed to be concerned in an essential way with moral questions. The theory can be perfectly successful from a scientific point of view even if it provides no moral guidance at all. In fact, we might find that the theory carves up the phenomena in a way that is completely orthogonal to the categories that prove most relevant in our moral thinking. But no matter. As long as the theory does well according to the distinctive criteria of science (empirical adequacy, simplicity, etc.), we are supposed to consider it a success.

We can now get a better handle on the question as to whether or not folk psychology is something like a scientific theory. In addressing this question, it is not enough just to ask whether or not folk psychology is sensitive in the right way to the kinds of considerations that play a role in scientific inquiry. We also need to know whether it resembles science in *excluding* the kinds of considerations that are usually excluded from scientific inquiries.

4. At this point, it might be thought that we really do have quite good reason to assume that folk psychology excludes the very same sorts of considerations that are normally excluded from scientific inquiries. After all, it is a conspicuous fact about our modern age that scientific approaches have proved extraordinarily successful in the systematic research programs where they are most commonly employed. One might therefore be tempted to conclude that the most effective way to proceed as folk psychologists would be to use almost exactly the same methods we find used in scientific inquiry.

But perhaps this conclusion is a bit too hasty. Clearly, there are some important differences between what we are looking for in a scientific research program and what we are looking for in a folk theory like folk psychology. So it is at least conceivable that the approach that best serves our needs in scientific research programs will not also best serve our needs in folk theories. Before we can determine whether or not there is reason to suspect that folk psychology uses a scientific approach, we therefore need to look in more detail at the advantages and disadvantages of that approach more generally.

One of the chief advantages of the scientific approach is its unparalleled predictive power. By excluding many of the criteria used in other kinds of inquiry, a scientific investigation can arrive at theories that do an extraordinarily good job at predicting the phenomena under study.

But this predictive power comes with a price. A scientific theory is a highly special-purpose tool. It might do an excellent job when our aim is to make predictions, but it won't necessarily prove helpful in all of the other tasks for which we ordinarily use complex conceptual thought. In particular, it won't necessarily carve up the phenomena in a way that proves helpful for making moral judgments.

Think, for example, of the various ways in which we might divide people up into categories. One approach would be to develop concepts that did the best possible job of predicting and explaining behavior. (And here we might end up with concepts like *person with high serotonin levels.*) But the categories we construct using this approach may turn out to be not ideal when it comes time to make moral judgments. Indeed, it may turn out that the categories that prove most helpful in making moral judgments are completely orthogonal to the categories that prove most helpful in generating predictions and explanations.

Assuming that we do want to make moral judgments, it seems that we will need to develop additional, non-scientific concepts that help us to pick out the morally relevant categories. Ultimately, we will then be left with two different ways of carving up the same class of phenomena. We will have concepts that pick out the categories that prove most helpful in prediction and explanation (e.g., person with high serotonin levels) and also concepts that pick out the categories that prove most helpful in making moral judgments (e.g., morally good person). We will then need a complex system of rules that enables us to move from one set of concepts to the other.

For cognitively limited creatures like ourselves, this level of specialization might be a major problem. We would have to retain in our minds two distinct systems of concepts, two distinct kinds of psychological mechanisms, two distinct sets of propositional attitudes. Whenever we were engaged in tasks that involved both prediction and moral judgment, we would have to shift back and forth from one system of categories to the other. All this would impose a substantial demand on our cognitive resources.

In short, the sort of approach we now associate with science has both advantages and disadvantages. The chief advantage lies in its *predictive power*; the chief disadvantage lies in the resulting *conceptual complexity*.

5. There is, however, another possible approach. Instead of having one system of concepts for use in generating predictions and then a second, completely separate system of concepts for use in making moral judgments, we could have a single system of concepts that was used for both of these tasks. This single system of concepts might not do a perfect job either at generating predictions or at making moral judgments, but it could do at least an adequate job of *both*. Hence, although this system of concepts might not afford us the greatest possible predictive power, it would do quite a bit to reduce the amount of cognitive complexity we needed to handle.

For an analogous case, consider the various ways we might come to conceptualize the weather. In thinking about the weather, there is a need to make *predictions* about what conditions will arise in the future, and there is also a need to make *evaluations* of whether these conditions are good or bad for certain purposes. What sorts of concepts would best enable us to achieve these goals? One approach would be to have a system of concepts that was specifically suited to the task of making predictions and then another, entirely separate system of concepts that was specifically suited to the task of evaluation. But such an approach might leave us with a large and unwieldy array of distinct ways of carving up the same class of phenomena. We might therefore be better served by a single system of concepts that wasn't ideally suited either for prediction or for evaluation but could serve us at least fairly well in both of these tasks.

It is certainly conceivable that folk psychology uses a system of concepts that works more or less along these lines. That is, it is conceivable that folk-psychological concepts are constructed in such a way that they do an adequate job at helping us both with prediction and with moral judgment, though perhaps without doing an absolutely ideal job in either of these two domains. What we want to know now is whether there are any general theoretical arguments against the view that folk psychology works in this way.

Thus far, we have been considering one possible argument. This argument relies on an analogy between folk psychology and systematic science. It points out that systematic scientific research programs typically *don't* try to develop a small set of concepts that enable us to do at least passably well at a wide variety of different tasks. Instead, they typically seek to develop concepts that enable us to do the best possible job at a specific range of tasks (prediction, explanation, etc.), even if they thereby end up coming up with concepts that aren't especially helpful in the task of making moral judgments. The argument then suggests that this fact about the concepts used in systematic science gives us reason to expect to find something similar in the concepts used in folk psychology.

At least for the sake of argument, we have been accepting all of the relevant claims about the nature of systematic science. The key question then becomes whether these claims can justify the relevant inferences about folk psychology.

6. But when the question is put in these terms, one notes immediately that folk theories are quite different from the sorts of theories one typically develops in systematic scientific research programs. Clearly, the two kinds of theories occupy two very different kinds of roles in our lives, and there is therefore little reason to expect that people look to them to fulfill the same needs. Most importantly for present purposes, it seems that people are far more reluctant to tolerate conceptual complexity in a folk theory than they are in the theories they employ in systematic research programs.

In systematic research programs, one can easily deal with the problem of conceptual complexity through a division of cognitive labor. No individual researcher needs to learn all of the scientific concepts; each only needs to know the concepts used in one particular domain of inquiry. Thus, science as a whole can acquire an extraordinary level of conceptual complexity even without any individual person grasping more than a tiny fraction of the total.

This solution is not available in the case of folk psychology. We cannot make do with a system in which one person only knows the emotion concepts, another only knows the trait concepts, and so on. We will only be able to do a tolerable job of getting around in the world if each person has some grasp of the whole of folk psychology. In fact, this seems to be one of the fundamental differences between folk theories and systematic research projects. We do not look to folk theories for a system that can serve, at least in principle, to generate perfectly accurate predictions. We look to them for tools that can help creatures like us — with all of our cognitive limitations — to accomplish certain practical goals.

Ultimately, then, it seems that we have good reason to expect that the concepts used in folk psychology will differ in certain respects from the concepts used in systematic research. In systematic research projects, one should expect to find an enormous array of different concepts, with each concept highly specialized for one particular use. But there is good reason to expect that folk theories will work somewhat differently. In a folk theory, one should expect to find concepts that are less highly specialized and can therefore be used in a wider variety of different tasks. Each concept might be specific to one particular domain of phenomena, but it will be constructed in such a way as to help us do almost anything we might want to do with the phenomena in that domain. Thus, instead of expecting to find a clear distinction between concepts used for prediction and concepts used for moral judgment, one should expect to find concepts that are not specialized for either of these two tasks but are constructed in such a way that they can do a decent job of both.

III

There seems to be a widely shared intuition that moral considerations just *couldn't* be playing any fundamental role in the basic concepts of folk psychology. Researchers who hold this intuition have not backed it up with systematic arguments. In fact, they have not even mentioned it explicitly. Yet the underlying intuition comes through quite clearly in the incredulous stares one receives whenever one suggests that some particular folk-psychological concept might be best understood as having moral features.

My concern here has been with the question as to whether there actually are any general theoretical arguments in favor of this intuitive view. I focused in particular on the argument that we have reason to expect that folk psychology will show certain fundamental similarities to scientific inquiry. This argument did not fare especially well on closer inspection. In fact, it seems that we actually have some reason to expect that folk psychology will differ from science in the relevant respects.

Of course, it is possible that there really are good arguments for the view that moral considerations can’t play any fundamental role in folk-psychological concepts and that these arguments have simply eluded my grasp thus far. In that case, I would want to know exactly what the relevant arguments are. Clearly, we should not reject a hypothesis simply because it goes against our philosophical preconceptions. What we need now are definite theoretical proposals that generate testable predictions about the structure of people’s folk-psychological concepts.

References

Adams, F. & Steadman, A. (2004). Intentional Action in Ordinary Language: Core Concept or Pragmatic Understanding? *Analysis,* 64, 173-181.

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

Adams, F. & Steadman, A. (forthcoming). Folk Concepts, Surveys, and Intentional Action. *Proceedings of the International Conference "Intentionality, Deliberation and Autonomy—The Action Theoretic Basis of Practical Philosophy,"* Sienna Italy.

Andrews, K. (2006). The Functions of Folk Psychology. Unpublished manuscript. York University.

Baron-Cohen, S. (1995), *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.

Bloom, P. (2004). *Descartes' Baby: How the Science of Child Development Explains What Makes us Human*. New York: Basic Books.

Bratman, M. (2000). Valuing and the Will. *Philosophical Perspectives: Action and Freedom*, 14, 249-265.

Churchland, P. (1981). Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy,* 78, 67-90.

Copp, D. (1995). *Morality, Normativity, and Society*. New York: Oxford University Press.

Faucher, L., Mallon, R., Nichols, S., Nazer, D., Ruby, A., Stich S. & Weinberg, J. (2002). The Baby in the Labcoat: Why Child Development Is An Inadequate Model for Understanding the Development of Science. In P. Carruthers, S. Stich & M. Siegal (eds.), *The Cognitive Basis of Science.* Cambridge: Cambridge University Press, 335-362.

Gopnik, A., C. Glymour, D. Sobel, L. Schulz, T. Kushnir, & D. Danks (forthcoming). A Theory of Causal Learning in Children: Causal Maps and Bayes-Nets. *Psychological Review.*

Gopnik, A. & A. N. Meltzoff (1997). *Words, Thoughts, and Theories*. Cambridge, Mass.: Bradford, MIT Press.

Gopnik, A. & Wellman, H. (1992). Why the Child's Theory of Mind Really *Is* a Theory. *Mind and Language*, 7, 145-171.

Harman, G. (forthcoming). Intending, Intention, Intent, Intentional Action, and Acting Intentionally: Comments on Knobe and Burra. *Journal of Cognition and Culture*.

Hutto, D. (2004). The Limits of Spectatorial Folk Psychology. *Mind and Language*. 19, 548-73.

Keil, F. (1989) *Concepts, Kinds, and Cognitive Development*. Cambridge, MIT Press.

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

Knobe, J. (2003b). Intentional Action in Folk Psychology: An Experimental Investigation. *Philosophical Psychology,* 16, 309-324.

Knobe, J. (2004). Intention, Intentional Action and Moral Considerations. *Analysis,* 64, 181-187.

Knobe, J. (2005a). Folk Psychology and Folk Morality: Response to Critics. *Journal of Theoretical and Philosophical Psychology*, 24, 252-258.

Knobe, J. (2005b). Theory of Mind and Moral Cognition: Exploring the Connections. *Trends in Cognitive Sciences*, 9, 357-359.

Knobe, J. (forthcoming). The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology. *Philosophical Studies*.

Knobe, J. and Burra, A. (forthcoming). Intention and Intentional Action: A Cross-Cultural Study *Journal of Culture and Cognition.*

Knobe, J. & Mendlow, G. (2004). The Good, the Bad and the Blameworthy: Understanding the Role of Evaluative Reasoning in Folk Psychology. *Journal of Theoretical and Philosophical Psychology*, 24, 252-258.

Leslie, A., Knobe, J. & Cohen, A. (forthcoming). Acting intentionally and the side-effect effect: 'Theory of mind' and moral judgment. *Psychological Science.*

Lewis, D. K. (1989). Dispositional Theories of Value. *Proceedings of the Aristotelian Society*, Supp. 63: 113-137.

Malle, B. F. (forthcoming). The relation between judgments of intentionality and morality. *Journal of Cognition and Culture.*

Malle, B. F., & Edmondson, E. (2006). What are Values? A Folk-Conceptual Investigation. Unpublished Manuscript. University of Oregon.

McCann, H. (2005). Intentional Action and Intending: Recent Empirical Studies. *Philosophical Psychology*, 18, 737-748.

McCauley, R. N. (2000). The Naturalness of Religion and the Unnaturalness of Science. In F. Keil and R. Wilson (eds.), *Explanation and Cognition*. Cambridge: MIT Press, 61-85.

Meeks, R. (2004). Unintentionally Biasing the Data: Reply to Knobe. *Journal of Theoretical and Philosophical Psychology*, 24, 220-223.

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

Morton, A. (2003). *The Importance of Being Understood: Folk Psychology as Ethics.* London: Routledge.

Morton, A. (forthcoming). But Are They Right? The Prospects for Empirical Conceptology. *Journal of Cognition and Culture.*

Nadelhoffer, T. (2004). Blame, Badness, and Intentional Action: A Reply to Knobe and Mendlow. *Journal of Theoretical and Philosophical Psychology,* 24, 259-269.

Nadelhoffer, T. (2005) Skill, Luck, Control, and Intentional Action. *Philosophical Psychology*, 18, 343-354.

Nadelhoffer, T. (forthcoming a). Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Jury Impartiality. *Philosophical Explorations*.

Nadelhoffer, T. (forthcoming b). On Saving the Simple View. *Mind and Language.*

Nichols, S. & Ulatowski, J. (2006). Intuitions and Individual Differences: The Knobe Effect Revisited. Unpublished manuscript. University of Utah.

Pinker, S. (1997). *How the Mind Works*. New York: WW Norton.

Premack, D. & Woodruff, G. (1978). Does the Chimpanzee Have a Theory of Mind? *Behavior and Brain Sciences,* 4, 515-526.

Sayre-McCord, G. & Smith, M. (2005). Desires... and Beliefs... of One's Own. Unpublished manuscript. University of North Carolina-Chapel Hill.

Sellars, W. (1956). Empiricism and the Philosophy of Mind. *Minnesota Studies in the Philosophy of Science*, 1, 253-329.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.

Stich, S. & Nichols, S. (1998). Theory Theory to the Max. *Mind and Language*, 13, 421-49.

Sverdlik, S. (2004). Intentionality and Moral Judgments in Commonsense Thought about Action. *Journal of Theoretical and Philosophical Psychology,* 24, 224-236.

Turner, J. (2004). Folk Intuitions, Asymmetry, and Intentional Side Effects. *Journal of Theoretical and Philosophical Psychology,* 24, 214-219.

Watson, G. (1975). Free Agency. *Journal of Philosophy*, 72, 205-20.

Watson, G. (1987). Free Action and Free Will. *Mind*, 96, 145-172.

Wilkes, K. (1981). Functionalism, Psychology, and the Philosophy of Mind. *Philosophical Topics*, 12, 1.

Young, L., Cushman, F., Adolphs, R., Tranel, D., & Hauser, M. (forthcoming). Does Emotion Mediate the Effect of an Action's Moral Status on its Intentional Status? Neuropsychological Evidence. *Journal of Cognition and Culture.*

Yoo, J. (2004). Folk Psychology and Moral Evaluation. *Journal of Theoretical and Philosophical Psychology*, 24, 237-251.