

Folk Psychology as a Model

Peter Godfrey-Smith

Philosophers' Imprint
<www.philosophersimprint.org/005006/>
Volume 5, No. 6
November 2005
©2005 *Peter Godfrey-Smith*

1. Introduction

I take *folk psychology* to be the basis—whatever it is—of our ability to describe, interpret, and predict each other by attributing beliefs, desires, hopes, feelings, and other familiar mental states. The nature of folk psychology has been the topic of extensive debate. Much of this debate has been structured by an opposition between the *theory-theory* and *simulationism*.¹ The theory-theory holds that our folk-psychological capacities involve the use of a theory, grasped by the interpreter, of how minds work. Simulationism holds that we interpret and predict others by simulating their thought processes in our own reasoning mechanisms. Recently, this standard opposition has been transforming. Perhaps the difference between the two options "collapses" under closer scrutiny (Davies and Stone 2001). Perhaps, alternatively, the original distinction does make sense, but the true view is a more complex one, involving elements from both of the standard stories and more besides (Nichols and Stich 2003).

This paper will present a different option. I argue that folk psychology should be seen as something like a *model*, in a specific sense of this term. This idea will be presented initially as a modification of the theory-theory, a modification that draws on ideas from recent philosophy of science. But once the main ideas are on the table, we will also see the possibility of a new kind of "mixed" view, with theory-like and simulationist elements in natural combination.

To what extent is the view developed here a novel one? Informal talk of a folk-psychological or belief-desire "model"

¹ Davies and Stone (1995), Stone and Davies (1996), and Caruthers and Smith (1996) are useful collections, containing classic and more recent papers on both sides of the issue.

Peter Godfrey-Smith is a Professor of Philosophy in the Research School of Social Sciences at the Australian National University and Visiting Professor of Philosophy at Harvard University.

is not unusual.² Generally, people using this term do not mean to mark any contrast between models and theories, however, a contrast that will be central to this paper. (The term 'model' is very often used to acknowledge a simplification, or a merely provisional commitment to a view.) A view of folk psychology that is close to mine can be found in a valuable recent paper by Heidi Maibom (2003). Maibom makes use of some of the same material from the philosophy of science that I will employ. We put this material to different uses, however (section 3, below). Peter Menzies (forthcoming) has also been developing a view of the same kind, drawing on a slightly different concept of 'model'. Looking further back, I see some aspects of this paper as an alternative development of themes in Dennett's "Three Kinds of Intentional Psychology" (1981).³

2. One sense of 'model'

Ideas about folk psychology have often been borrowed from philosophy of science. This seems natural: if folk-psychological skill involves a kind of theoretical understanding, we might look at accounts of how theories work elsewhere, perhaps how they work in science. There are obviously two possible sources of error here. First, folk psychology might be very different from science, even if both involve the use of theory-like constructions. Second, even if science and folk psychology are similar, *philosophy* of science

² Explicit talk of models as opposed to theories has appeared, in particular, in some recent action theory, as will be discussed in the final section below. This tradition stems at least in part from Grice (1974-75), who introduced a methodology of "creature construction", in which a hierarchy of psychologically simpler and more complex agents is described. I should note that it is not the mere *introduction* of sets of simplified imaginary agents that is relevant here, but the claim that folk-psychological concepts like *belief* can only, or can best, be understood in terms of the features of such model systems. Grice does make a claim of this kind.

³ A rougher sketch of the view in this paper can be found in Godfrey-Smith (2004), which includes more detailed discussion of Dennett's ideas.

might not be doing a good job at describing how theorizing works in science. I will begin by looking at this second possibility.

Here are some ideas in the philosophy of science that have influenced many lines of argument in philosophy, but especially work in the philosophy of mind.⁴

1. There is basically one way that theories operate in different parts of science.
2. Theorizing is largely a search for generalizations, or perhaps for "laws". Some of these generalizations are empirical or "phenomenological". Others are cast in terms of posited unobservable or "theoretical" entities and properties. (Philosophically, there are several rival ways of handling the apparent role of unobservable entities here; those debates are orthogonal to the present discussion.)
3. Theories have truth conditions (or some closely related kind of semantic content). Theories make commitments about how the world is.

Here is an alternative view of some of the same issues in the philosophy of science.

- 1*. There is a variety of approaches to theorizing in science. One of these approaches is model-building.
- 2*. A model-builder's usual goal is to construct and describe various hypothetical structures.

⁴ I see Fodor's work (1981, 1987) as clearly exemplifying these assumptions. Braddon-Mitchell and Jackson (1996) is another good example, as is Crane (2003). In psychology, Gopnik and Meltzoff (1987) explicitly accept a view that is at least close to this, although they emphasize causal mechanisms more than the philosophers do. Maibom (2003) gives other relevant citations.

These structures are used to help us understand some actual target system or systems. Generally, the understanding is supposed to be achieved via a *resemblance* relation between the hypothetical and real systems. But both the degree and kind of resemblance that is sought are adjustable.

- 3*. So a model itself does not contain commitments about what the target system is like, or even which system is the target. A model and its application are two different things.

In understanding this kind of model-based science, a good starting place is Ron Giere's work (1988, 1999). The account I use here is not exactly Giere's. Giere, like many others, is too inclined to see the model-based view as a new and superior account of *all* theorizing in science. In this respect, Giere's view is part of the tradition known as the "semantic view of theories" (Suppes 1960, Suppe 1977). Writers in and around that tradition have sometimes criticized Giere's version of the view, especially for its reliance on an informal and vague notion of resemblance between model and target (Chakravarty 2001). But Giere's account can most usefully be seen as a contribution to a different project from the one associated with the semantic view, and this project is one in which the "slippery" features of resemblance relations are quite important. According to this alternative view, model-building is one, but not the only, kind of theorizing found in science. Model-building is a rather distinctive scientific strategy, and a good description of it will not generalize to other kinds of scientific work (see also Wimsatt, 1987; Weisberg, forthcoming; Godfrey-Smith forthcoming[a]). Within this kind of science, the ability to describe and develop model systems in some detail, while remaining cautious or

flexible about the particular respects in which the model might resemble the target system, is an essential tool. Modeling is especially useful when our knowledge of the target system is poor, and its workings are complex.⁵

In describing model-based science, it is simplest to start by thinking of a model as a single hypothetical structure that can be used to represent a single target system. The actual situation is usually more complicated. What is referred to as "a" model is usually a class of hypothetical systems, similar to each other in general pattern, and constructed from a common repertoire of elements. When a scientist has facility with a model, the scientist has an understanding of a whole category of hypothetical systems. The scientist can describe the particular behavior of specific systems in this category, and (hopefully) has a general sense of how these systems work—what depends on what, why they do what they do.

Mathematical models of evolution by natural selection provide an example of this kind of work. Our modern understanding of evolution is organized, in large part, by a family of abstract models (Roughgarden 1979, Lloyd 1988, Rice 2004). This family includes things like "the one-locus model" of natural selection. But "the" one-locus model includes a large class of idealized evolutionary scenarios. These scenarios are constructed from the same ingredients (genotypes, fitness values, mating assumptions, etc.), and the ingredients are combined in each case according to a common pattern. A scientist who is proficient with the model can construct specific cases of the one-locus model that are of empirical or theoretical interest, predict how they will behave, and compare these cases to others.

In many examples of model-based scientific work, a distinction can be made between more "top-down" and more

⁵ In Godfrey-Smith (forthcoming[a]), the contrasts between the project of the "semantic view of theories" and the analysis of a strategy of model-based science are examined in more detail.

"bottom-up" approaches. A modeler may begin with a general structural principle, such as an equation or system of equations, and then look at the consequences of filling in particular values to terms in this mathematical structure. This is, in a sense, a top-down approach. The other, more bottom-up approach is to begin with a set of hypothesized elements or ingredients and work up by exploring their possible interactions. An explicit general principle such as an equation might emerge from this analysis, but it need not do so. This distinction between more top-down and more bottom-up strategies will be relevant later in this paper.

The exact referent of the term 'model' is often ambiguous in this kind of science. Sometimes the term will refer to a specific hypothetical system, sometimes to a class of these systems, sometimes to the conceptual toolkit that is used to describe them.

Before moving on to folk psychology, I pause to emphasize the features of model-based theorizing that I will be putting to particular use. Two scientists can use the same model to help with the same target system while having quite different views of *how* the model might be representing the target system. I will describe this situation by saying that the two scientists have different *construals* of the model. Although there are different possible ways of carving things up here, I treat the choice of a target system as distinct from the construal of the model. The construal has to do with *how* the model is taken to represent the target.

For example, one scientist might regard some model simply as an input-output device, as a predictive tool. Another might regard the same model as a faithful map of the inner workings of the target system. So both scientists, in a sense, are hoping for a resemblance between model and target, but they are looking for very different *kinds* of resemblance. Between these extremes there are obviously many

intermediate kinds of resemblance that might be sought; another scientist might see the model as resembling the target with respect to basic architectural features but not with respect to details.

Two scientists might also use the same model to understand different target systems. And a scientist might change his or her use of the model over time. Theories, as they are usually understood by philosophers, make claims about the world; theories make commitments about how things are, usually in the form of generalizations. Models, in my sense, do not themselves say anything about the world. Models are structures that can be *used* by scientists to say various different things about the world, by means of commentaries that accompany models but are distinct from them. (Giere calls these "theoretical hypotheses".) But the role of these commentaries is very subtle. For example, models can be used to make *usefully vague* claims about a target system. In the model-building tradition in science, claims are often expressed in the form: "what is going on here is something like this...", where what follows is the specification of a hypothetical system. A model can also be used to study a target system while remaining studiously neutral on many questions about the appropriate construal. Lastly, a model can be explored and elaborated without reference to its possible application to real systems at all; it can be studied as a structure in its own right.

3. Folk psychology as a model

My suggestion is that one aspect of ordinary folk-psychological skill might best be described not as grasp of a theory but as something like *facility with a model*. Perhaps when we find ourselves engaging in ordinary, unreflective folk-psychological interpretation, we are bringing something like a model to bear on the person we are trying to interpret.

Basic facility with the folk-psychological model does not require using a particular construal of it. Many construals are possible. And it is also possible to have facility with the model, and have a sense of which target systems are appropriate for it, while not having much of a construal at all.

That is the main idea in this paper, and the rest of the discussion will be concerned with spelling it out in more detail, qualifying it where necessary, and showing how it casts new light on some otherwise puzzling debates.

I should emphasize that at many points in the story I am uncertain about how best to develop the details. I do not want to claim that folk psychology is *exactly* like a theoretical model of the kind used in science, and it is unclear which features of scientific models might have analogies in the folk-psychological case. The similarity between these two kinds of theoretical understanding might be loose or tight. Some of my points can be made with the aid of only a very partial analogy between the two phenomena.⁶

We should not rule out the possibility, however, of a stronger resemblance between the scientific and the folk activities. It is possible that a basic part of our psychological apparatus is a facility for what we might call *model-based understanding*. This skill involves the imaginary construction of simplified structures for the purpose of understanding more complex systems. This might be a skill seen in an elaborate form in some kinds of science, and in a more rudimentary form in folk theorizing. The idea of a psychological capacity of roughly this kind, and its possible role within scientific

⁶ Although folk psychology is not itself a scientific enterprise, and hence is likely to be only similar to scientific model-building, there is a related set of ideas that *are* part of science and that can be given a more literal treatment in the terms employed in this paper. In Godfrey-Smith (forthcoming[b]) I argue that the commitment to *mental representation* seen in contemporary psychology should be seen as a case of model-based science. This has consequences for philosophical attempts to analyze the commitments of the "representational theory of mind".

thinking, has been investigated by a number of workers within psychology and science studies (Johnson-Laird, 1983; Nersessian, 1999; Gentner 2002).⁷ If there is such an ability, the products of some kinds of folk theorizing might have important features in common with models in the scientific sense, and *both* of these may contrast with some more traditional philosophical notions of theory. Further, once we recognize the possibility of model-based understanding—in this sense—as a distinctive psychological capacity that operates both inside and outside of scientific contexts, we can note that folk psychology could have model-like features that are not very science-like.⁸ Here, however, I will proceed by emphasizing the analogy with scientific models.

Let us now look more closely at how folk-psychological interpretation might be understood in terms of the use of a model. I suggest that the similarities are vivid and immediate. To the extent that folk-psychological skill looks like a theoretical matter, it looks more like model-building than theorizing as traditionally understood. And as we will see below, the model-based view also has the capacity to include elements that are usually seen as antithetical to the theory-theory approach.

An ordinary folk-psychological interpreter seems to exhibit a collection of tools and skills that are quite similar to those seen in the scientific case. The interpreter, like a scientist, has an understanding of a general structure or schematic pattern that can have many specific instantiations. In the folk-psychological case, this is the familiar basic picture of the contents of normal minds (beliefs, memories, hopes,

⁷ A number of senses of 'model' have been used within this tradition in psychology, and some of them have no particular similarity to the sense I use here. Sometimes, for the psychologists, any representation is referred to as a mental model. But there is real continuity between key parts of this work and the development of relevant ideas in the philosophy of science, as has been emphasized by Nersessian.

⁸ I am indebted to John Hughes for emphasizing this point.

fears...) and how these contents typically interact. The interpreter, like the scientist, also has the ability to construct *specific* hypothetical systems to deal with particular empirical cases. Folk-psychological interpreters can rapidly put together specific, filled-out psychological profiles, to explain and predict the actions of individual agents. Some of these specific psychological profiles are extremely fragmentary and minimal, while others are rich and detailed. They are put together out of a repertoire of standard elements, including states like beliefs and preferences that take that-clauses and are linked by logical relations, along with such things as emotions, sensations, and moods.

So suppose a discussion begins between two people, about the thinking and decision-making of someone else. One speaker introduces the project of giving a folk-psychological interpretation of the subject. What follows is the sketching of different ways of filling in a specific psychological profile for the person in question—specific hopes, fears, and plans. If the aim of the discussion is prediction of the third party's likely actions, then the point of the modeling is to generate an "output" in the form of a predicted decision or action. Alternatively, the aim may be to construct a plausible etiology for an action that is known to have occurred.

In either case, basic folk-psychological interpretation consists in constructing a *model psychological profile* for the person in question. A lot of interpretive and predictive moves can be made with the model before it becomes necessary to wonder about exactly *how* the model corresponds to the target—what sort of resemblance is envisaged. But as the discussion goes on, it might eventually become necessary to establish at least some constraints on how the model is being construed. This, I am suggesting, is a distinct move, and one that is usually optional. The role of construals is

discussed extensively in the next section.

With these main ideas on the table, I will discuss the paper by Maibom mentioned in the Introduction (2003). Maibom draws on some of the same tools from the philosophy of science, such as Giere's treatment of models (1988), but we put this material to different uses. First, Maibom's main emphasis is on how it is possible for ordinary people to have *knowledge* of folk psychology when they are not able to produce explicit and non-trivial generalizations, as required by the traditional philosophical picture of theories. The inability of people to cite useful folk-psychological generalizations has been taken to motivate an appeal to tacit knowledge, of the kind associated with the Chomskian program in psycholinguistics. Maibom argues that if we see folk psychology as organized around models rather than generalizations, we can regard folk-psychological knowledge as ordinary, everyday knowledge rather than the special tacit kind. Maibom also uses a Giere-style account of models to treat the role of idealizations in folk psychology.

I agree with Maibom's critical points. In particular, she has removed one argument for an appeal to tacit knowledge. Once this argument has been set aside, it is an open question whether or not tacit knowledge might be invoked in this case, however. The account of folk psychology I present here is neutral on this question.

Secondly, there is a key difference in our application of the ideas we draw from the philosophy of science. Maibom and I both use Giere's work, but she allies it to the work of Suppes and the "semantic view of theories". That tradition aims at a wholly general account of scientific theorizing, and it uses strict, formal concepts like isomorphism, rather than a more flexible and possibly holistic notion of similarity, to treat the relations between models and empirical systems. In contrast, I treat Giere's picture as a good initial description

of a specific kind of scientific work. Work of this kind is often useful when scientists are faced with problems of complexity and ignorance, and this scientific strategy makes use of a flexible and context-sensitive notion of relevant similarity between model and target. That is the feature of model-based science that permits a single model to be employed with various different construals, as I call them. This distinctive kind of flexibility is not discussed by proponents of the "semantic" view. In that literature, the vagueness associated with similarity relations appears as a problem rather than as a positive feature. Accordingly, Maibom's treatment of folk psychology does not include what I treat here as a central advantage of a model-based view, the role played by flexibility in how a single model can be construed.⁹

The next step in my presentation is a discussion of relations between the view presented here and the "simulationist" account of folk psychology (Stone and Davies 1996). This comparison is important because simulationism is often expressed using the idea of a model: folk psychology is explained in terms of the use of one's own psychological machinery as a model of others.¹⁰ How does this idea relate to my proposal?

I will make two points. The first is that the simulationists use a different sense of 'model' from the sense used in this paper. The basic version of my proposal is within the theory-theory camp. But the second point is that there is also the possibility of a mixed version of my view, with elements of simulationism and the theory-theory combined.

First, the usual sense of 'model' seen in simulationism is

⁹ Maibom's treatment may also not include the possibility of a "mixed" theoretical/simulationist view as outlined below in this section.

¹⁰ I should note that Robert Gordon, one of the original proponents of simulationism, explicitly rejects this common way of talking about the simulationist option (Gordon 1995, section 5). Jane Heal (1986) does not talk explicitly of models, but her "replicative strategy" can fairly be described, I think, as treating oneself as a physical model of others.

not my sense. Terminological issues are problematic here, but I will mark the distinction by talking of *theoretical* and *physical* models. A *theoretical* model is the kind I have been discussing so far. A theoretical model is a hypothetical system (or family of such systems), specified using some sort of representational medium, that is constructed for the purposes of comparison to a target. A physical model is an actual physical system that is used to represent another, again via a resemblance between the two. So a physical model need not be imagined or represented; it is a physical object.

A scaled-down wing in a wind tunnel is an example of a physical model, though not all physical models need have that degree of similarity to their targets. A bacterial population evolving in a petri-dish, created to investigate some general features of natural selection, is also operating as a physical model. An imaginary bacterial population specified using equations written on paper is a theoretical model.

Simulationism claims that folk-psychological prediction works by one person's using his *own* reasoning mechanisms as a *physical* model of another person's. My hypothesis is that one person predicts another by using a *theoretical* model—something quite distinct from the interpreter's own reasoning mechanisms. Part of the point of simulationism involves a certain kind of *economy* in the picture of the mind being offered. One's own reasoning mechanisms are seen as co-opted for use in interpreting others: why do we need a theory when we can use ourselves as physical models? My hypothesis, in its basic form, does not have this feature.

While I treat the distinction between the theory-theory and simulationism as real, there is also the possibility of a mixed view. In fact, the model-based approach suggests a way to combine the two kinds of mechanisms in a natural way, without dissolving the distinction between them.

In section 2, above, I distinguished between two ap-

proaches to constructing model systems, a more "top-down" approach and a more "bottom-up" one. The top-down approach begins with a general principle or pattern and then generates specific cases from it. This will include procedures like setting specific values to parameters in equations. The bottom-up approach is to begin with a set of specific hypothesized elements and explore their possible interactions. This can be done in cases where the theorist does not have a general principle or summary equation available. This second kind of work is associated especially with computer simulations as opposed to "analytical" techniques.

The reference just made to computer simulations might suggest the analogy I want to draw. A folk-psychological modeler might in some cases have access to a plausible set of folk-psychological elements with which to build a model of some agent, but insufficient general knowledge to determine the likely relations between these elements. Such a modeler might then use other ways of working out how these elements might interact. One way to do this would be to employ the simulationist "off-line" operation of the interpreter's ordinary reasoning mechanisms.

The use of one's own reasoning mechanisms to generate a picture of how a set of psychological elements might interact is an extreme case of the more general phenomenon of bottom-up modeling. Usually, a modeler employing a bottom-up strategy will have a sense of the likely local interactions between elements in a model system; the aim is to generate an overall picture by working up from reasonable assumptions about the lower-level interactions. In the folk-psychology case, if an interpreter uses his or her own reasoning mechanisms as a physical simulator of how a particular set of hypothetical beliefs and desires might interact, the role for prior general knowledge is very minimal indeed.

The most important point here is that this bottom-up ap-

proach could be used in *conjunction* with the more theoretical, top-down approach; the two could work together. These would be two distinct ways of constructing folk-psychological model systems or pieces of them. Folk-psychological understanding in a particular case might be achieved with a model that contains *fragments* of both kinds—fragments contributed by knowledge of general patterns, and fragments contributed by off-line simulation. Such mixed views may have considerable promise. Simulationist ideas have been in the ascendant recently in psychology, but as Rebecca Saxe (2005) argues, there are both new and old data that speak strongly in favor of a substantial theoretical component to our folk-psychological capacities.

Once we see the possibility of these different styles of folk-psychological model-building, we also see the possibility of a reply to a simple but nagging objection to all theory-based views of folk-psychological concepts. The theory-theory tradition holds that our concepts of psychological states are, fundamentally, concepts of theoretical entities that are posited for explanatory purposes. Such a view has always looked awkward when applied to mental states that have a vivid first-person aspect to them, such as sensory states and the "basic" emotions such as anger and fear. Surely we do not need to introduce the idea of anger as a theoretical posit; we know it more intimately than that.

Philosophers have sometimes sought to fend off such worries by appealing to general views about how we could come to talk meaningfully about psychological states. The view defended in this paper has a far easier time with this issue.¹¹ My view can allow that elements might be *introduced* to the folk-psychological model via a number of different paths. Perhaps, to consider an extreme possibility, the mental state of fear is one that we can simply experience, la-

¹¹ I am indebted to Kim Sterelny and Dominic Murphy for discussion on this point.

bel, and then deploy when interpreting others. Fear thereby becomes one standard ingredient in a person's folk-psychological model-building. But even if this is how an interpreter becomes *acquainted* with fear, once it has been introduced to the model this element will acquire connections to a large range of other elements. It becomes part of a richer structure, much of which has far more tenuous relations to first-person experience.

The rest of the paper applies this view of folk psychology to various philosophical issues. A range of more empirical questions will not be addressed further. In particular, the view presented here is an hypothesis about *normal adult competence* in the domain of folk psychology; I have not defended views about where the model comes from. The ideas outlined here raise a number of questions about the development of folk-psychological skills in childhood. My distinction between basic *facility* with the folk-psychological model, and different *construals* of it, raises the possibility that children acquire the facility before they acquire any construal. But Alison Gopnik (personal communication) thinks that the empirical evidence does not support such an idea, and Maibom (personal communication) rightly insists that a model-based view of folk psychology is compatible with a range of different hypotheses about individual development. The model-based view also has interesting connections to empirical questions about the cultural variability of folk psychology (Nisbett 2003). Those issues, too, are left for another day. My aim in the remainder of the paper is to use the model-based view to try to cast new light on some familiar philosophical problems.

4. Versions and construals

A central feature of the account offered here is a distinction between the folk-psychological model itself and various pos-

sible construals of the model. In the previous section I suggested that everyday folk-psychological interpretation can proceed without a construal of the model being on the table at all, and that construals are in a sense secondary. But what are some examples of these different construals, and what roles do they play? I will approach this question by revisiting some famous parts of the philosophical literature, especially the eliminativism debates of the 1980s and 1990s.

People have argued for decades about what sort of claims folk psychology makes about the organization of the machinery inside our heads. After all, if folk psychology is a theory, then it must make claims about what the world is like. So there is a literature discussing what folk psychology is "committed" to, and hence whether it is likely to turn out to be false. Positions here range from the view that folk psychology is committed *no* claims about the character of the causal processes inside our heads (Dennett 1978), through fairly *weak* claims (Jackson and Pettit 1990),¹² *strong-ish* claims (Fodor 1987), to *quite strong* claims (Stich 1983, Ramsey, Stich and Garron 1991). Roughly speaking, I suggest that no one is right here. These philosophical views correspond to some sharply different construals of the basic folk-psychological model. Roughly, they fall on a scale from more instrumentalist to more realist construals. The model can be seen merely as an input-output device. The model can also be seen as a rough and minimal map of inner causal structure; its basic elements and the interactions between them can be seen as corresponding very roughly to real psychological structures. The model can also be seen as a much more detailed map of how psychological processes work.

¹² In discussion, Jackson insisted that these claims are in some respects quite *strong*, though they are also very epistemically safe. Behavioral data are rich enough to support quite strong inferences in this area. But the claims Jackson has in mind are still much weaker than the claims associated with folk psychology by people like Fodor and Stich.

Fewer features or more features of the model can be seen as having distinct, well-demarcated analogues in target systems. All these construals are *available* but none is *mandatory*. The folk-psychological model does not dictate its own construal. If we ask "What is folk psychology itself committed to?", the answer is "Nothing."

I argued that some standard options in the philosophical literature correspond to different construals of the model. But I do not want to suggest that construals of this kind are solely the creatures of philosophical reflection and discussion. Some attention to differences in construal seems to be part of everyday folk-psychological skill, at least in adult users. Everyday folk-psychological interpretation displays a good deal of flexibility with respect to which kinds of understanding it is used to achieve.

Contrast folk psychology on the freeway and in the lawcourt. In a lawcourt, folk psychology is used in the investigation of (among other things) *why* people did what they did. Special attention is paid to inner states—to motives, expectations, and fears. Did he fear for his life when he fired the fatal shot? Here folk psychology is used to try to make deep claims about inner causal processes (whether this activity is justified or not). On the freeway, in contrast, the main concern is just to get predictions about how others will behave. (Does he want to turn left there?) Folk psychology on the freeway is generally not concerned with what exactly is going on inside a target's head; folk psychology in the lawcourt often is.

In general, contexts in which folk psychology is used in the attribution of responsibility might tend to bring with them strong construals, and contexts in which its use is solely predictive will allow weaker ones. Some explanatory contexts might be intermediate between these two. I do not want to suggest that this introduction of a construal has to

be a definite, explicit thing. In a discussion of someone's actions and motives, initial moves might be made with a very minimal construal or no construal at all. But the speakers might find themselves gradually moving to a stronger or more definite construal, especially as they start to ask why-questions.

The role of construals is also important in understanding the application of the folk-psychological model to a wide variety of target systems. The most usual applications are to our fellow humans, of course, but the model can also be applied to non-human animals and to corporate entities like nations and political factions. So the same *model* can be applied to your neighbor, to his dog, and perhaps to his sector of society too. But it will generally be natural to construe the model somewhat differently in these cases.

So far I have presented the picture as if there was a single folk-psychological model. Or more precisely, I have supposed there is a single set of folk-psychological elements and one or a few ways of combining these elements into specific psychological profiles. The contrast has been between singularity of model and multiplicity of construals. But this is a simplification. It would be more accurate to say there is a single *core* folk-psychological model, plus a range of variants and elaborations. The core can be used alone, or it can be used as the foundation for more complex structures. This structural variation can, however, still be distinguished from variation in construals.

The core model includes such features as a distinction between beliefs and desires or preferences, the idea of sensory input and behavioral output, and the characteristic dependence of action on perceptions, memories, goals, and temptations. But it is easy to introduce more detail, and this can be done in a range of different ways.

The most useful example here is probably degrees of be-

lief. These are not found in the simplest versions of folk psychology, which treat belief as a binary matter. But degrees of belief can easily be introduced in an informal way ("how sure was she?"). This introduction of a quantitative concept of belief is quite normal within everyday contexts. But degrees of belief can then, of course, be used to develop the precise and elaborate versions of the model found in Bayesianism.

So the quantitative conception of belief is a variation on the basic model that is part of ordinary adult folk-psychological competence, but is developed in more detail in philosophical and psychological theories.¹³ Others variations are similar.

I have described two different dimensions of variation surrounding the basic folk-psychological model: we have different versions and different construals. Variation in versions is variation in internal, structural features of the model; variation in construals concerns how the model is used in relation to a target system. There is no need to get distracted by the question of whether the term "folk psychology" should refer to the core model itself, to the core model and a set of normally-used construals, to a larger family of models, or something else. All of these things are parts of the story.

At this point I should make explicit a feature of the present proposal that may have been clear already. Many normal, everyday folk-psychological ascriptions do not have definite truth conditions. Models, I have been emphasizing, do not themselves say anything about a target system. And

¹³ Compare the brief discussion of these issues in Lewis (1994). Lewis says that folk psychology is a tacitly known theory, and Bayesian theories of rational action and learning are "severely idealized versions" of parts of the theory (p. 321). Lewis adds that folk psychology itself "supplies the grains of salt" to be applied to these idealizations, and that the theory also sometimes supplies contrary idealizations (qualitative and quantitative notions of belief). I suggest that this is asking a great deal of a body of "tacit" knowledge that is supposed to be akin to our knowledge of grammar.

a model can be brought to bear on a target system without a definite construal being introduced. A lot of folk-psychological discussion, I suggest, involves the simple deployment of the folk-psychological model, without a construal being on the table. In fact, there is also typically a lot of indeterminacy in which *version* of folk psychology is being deployed. The model can be introduced into discourse and used productively despite this. And as the discussion goes on, there can sometimes be a sharpening-up.

For example, consider the US marine who sought conscientious objector status in early April of 2003. He said he wanted to leave the marines because he did not want to kill people. We ask: what was he *thinking*, when he joined the marines? (He joined, I understand, in the fall of 2002, when the Iraq war was brewing up.) When we ask this, we introduce the folk-psychological model into the discussion. But we don't necessarily introduce a specific version of the model; we don't necessarily ask "what were his degrees of belief?" as opposed to "what were his (qualitative) beliefs?" We can ask the question and make some progress on it while remaining vague about which exact version of the model and which construal is relevant. But we might, if the discussion goes on, introduce either a more elaborate version of the model, a more specific construal, or both. It is not just that we can come to look for a more *detailed* psychological profile; we may come to *use* the profile to ask a different kind of explanatory question from before. We might end up asking a hyper-realist question: what on earth was *going through his head*? But that need not be exactly the question we started with.

In response to an earlier version of these ideas, Daniel Stoljar (personal communication) insisted that I distinguish between multiple *uses* of folk psychology and the more radical hypotheses of ambiguous truth conditions or indetermi-

nacy. Various sorts of pluralism about folk psychology can be developed in a way that does not reject the idea that belief-ascriptions and the like have definite semantic contents. A speaker's use of a theory or sentence is not rigidly governed by the normal semantic properties of the theory or sentence. Stoljar's point is correct, and it introduces an entire different class of options. These are options where the semantic properties of folk-psychological interpretations are in good correspondence with *one* kind of application of folk psychology, and other applications are given some kind of special analysis. If the real standards for ascribing beliefs are high standards, then something must be said about why it is so often useful and socially unobjectionable to ascribe beliefs in the absence of any evidence that these standards are met. If the real standards are very low ones, then something must be said about why there are special contexts in which much higher standards seem to come into play. Such stories can certainly be told, and I reject "onus of proof" arguments in either direction. But I do insist that there is some reason to consider the option I am presenting here—an option that is semantically unconventional but takes at face value the diversity in how folk-psychological interpretations are employed.

I will illustrate this idea by describing an interesting moment in a philosophical discussion, at a conference for Philip Pettit at the Australian National University in July 2003. The discussion, which took place after Tim Scanlon's paper, was delving deeper and deeper into some "rationality constraints" on belief possession, having to do with the responsibility undertaken by any believing agent to recognize the role of reasons, and so on. Frank Jackson asked a question, and he began by saying that he is "in the faction that treats it as obvious that some [non-human] animals have beliefs." What followed, throughout the room, was a kind of

uncertain shifting of gears. Some people in the audience might have immediately mentally located themselves in the faction that denies that any non-human animals have beliefs. But what was more common, I think, was a kind of mental shift in how people were construing the folk-psychological model. "Right; sometimes it seems entirely reasonable to attribute beliefs to dogs and cats. I was thinking about a different kind of interpretation...."

What was happening here? Were people switching between two discrete "senses" of the term "belief"? If so, is there *another* sense they use with their three year-old child? How long is the list of senses? Alternatively, were they mentally signing up with one faction or the other but resolving to tell a special pragmatic story about the contexts in which a different kind of interpretation feels natural? I suppose those in Jackson's faction need not do much of this; they have the same lower standards for belief attribution in all cases. But those who do take seriously the recondite and subtle rationality constraints on interpretation found in many philosophical discussions would have to tell some kind of special story about the naturalness and utility of belief-ascription in contexts where those rationality constraints do not apply.

In this case, and various others discussed in this section, I suggest that the most natural view of what is going on is the two-layered one that distinguishes between models and their construals. There is a core structure employed in folk-psychological interpretation and discourse, a structure that can be employed without reflection, and without much agreement between interlocutors about *how* exactly it relates to the target system in question. There is also a range of ways in which the model can be construed, a range that encompasses various explicit choices, and vague inclinations, concerning the kind of understanding that the model is being used to attain.

5. Folk psychology in cognitive science and analytic philosophy

I conclude with a section that sketches some consequences of my position for discussions of folk psychology within (i) cognitive science and (ii) philosophical debates outside philosophy of mind that draw on folk-psychological concepts.

There has been much discussion of the question of which developments in cognitive science might vindicate or undermine folk psychology. I argued earlier that folk psychology itself does not make commitments, either weak or strong, about the mind. More positively, I suggest that the way to think about the situation is something like this. Cognitive science encounters the folk-psychological model as a starting point and as a source of structural ideas. Much of cognitive science itself proceeds by model-building, as opposed to other kinds of theorizing. As any scientific field changes, earlier models are "mined" for features to use in later ones. And this mining can proceed in a piecemeal way. Some parts of the model can be retained, perhaps transformed, while others are dropped.

Consider the cases of belief and desire. As cognitive science develops, models are developed that include states with a variety of resemblance relations—very close ones and much more tenuous ones—to folk-psychological beliefs and desires. For example, a *contrast* between beliefs and desires might be retained in a model, even when much else is not. Kim Sterelny's recent work, for example, includes an investigation of the evolution of desire (1999, 2003). Why should the capacity to represent goal states evolve, when it seems that at least many animals could get by with just non-representational *drives*? This is a good question. When Sterelny discusses the "descent of desire", however, the only thing that remains from the folk-psychological concept of desire is the *contrast* between "how things are" states (of pretty much any kind) and "what I want" states. The rest of

the baggage that might come with folk-psychological desires is simply ignored.

The concept of belief can be given the same treatment. A cognitive-scientific model might reject the idea of propositional content, in favor of something more holistic; it might reject the idea that all beliefs are available to all processing, in favor of a "modular" option; it might reject standard logic and probability theory as a theory of reasoning processes. Still, a contrast between information-bearing, "how things are" structures and structures expressing goals or preferred outcomes might be retained. One structural feature of the folk-psychological treatment of belief is kept while much else is abandoned. Does this cognitive-scientific model vindicate or undermine folk psychology? The question is misposed in *two* respects. First, the cognitive-scientific model, considered as a model and not combined with specific theoretical hypotheses, does not make definite commitments about what the world is like. But more importantly, what we have here is a very partial, but non-trivial, structural resemblance between an old model and a newer model. In situations like this, the vocabulary of folk psychology will be retained or rejected in scientific discussion in response to a range of considerations, many of them pragmatic. Does the speaker want to emphasize the continuity between this new model and folk psychology, or not? Does the speaker want to highlight, or background, the retention of some folk-psychological structure?

Lastly, I will make a comment about other parts of philosophy that draw on philosophy of mind and the debates about folk psychology. Here I especially have in mind fields like action theory, meta-ethics, moral psychology, and parts of epistemology. I suggest that philosophical literatures of this kind often engage in an kind of exploration-cum-elaboration of the folk-psychological model. "Elaboration"

here is meant to imply making *additions* to the model. Philosophers often present their work as investigations of the "structure of our concept" of thought, reasoning, deliberation, or motivation. I suggest that what they are really doing is elaborating, cleaning-up, trimming and regimenting the model—constructing *new* versions of it. They are not revealing pre-existing structure, structure that we all "grasp" but have not yet made explicit.

Analytic philosophers, however, sometimes seem reluctant to recognize the "additive" side of this kind of work—the refinement of, and addition of features to, a simpler pre-existing structure. This may be because it can then become unclear what these added features are *answerable* to. In some recent work in action theory, a field where particularly elaborate folk-psychological constructions are discussed, one kind of model-oriented methodology has been explicitly employed (Bratman 2000). This work adopts Grice's project of "creature construction" (1974-75), in which a hierarchy of simpler and more complex hypothetical agents is described. This move is not accompanied by a general treatment of models and their role, however. What consequences would it have if some of the rather baroque psychological structures discussed in action theory (extended hierarchies of higher-order attitudes, for example) were explicitly understood as "model-bound" entities, in a sense analogous to the scientific case? This would raise a number of new possibilities for thinking about the intended relation between these structures and the features of real-world agents. Here and elsewhere, I suggest that treating folk psychology as a model, or something very similar to one, is a step in the right direction.¹⁴

¹⁴ I am grateful to Karen Bennett, Daniel Dennett, Alison Gopnik, Marc Hauser, John Hughes, Frank Jackson, Doug Lavin, Heidi Maibom, Rebecca Saxe, Kim Sterelny, Daniel Stoljar, Tony Stone, and Michael Weisberg for discussions and correspondence. Two anonymous referees made useful criticisms. I am also grateful to audiences at Stanford, the Australian National University, Monash, and Wesleyan University for helpful comments.

References

- Braddon-Mitchell, D., and F. Jackson (1996). *The Philosophy of Mind and Cognition*. Oxford: Blackwell.
- Bratman, M. (2000). "Valuing and the Will." *Philosophical Perspectives* 14: 249-65.
- Carruthers, P., and P. Smith (eds.) (1996). *Theories of Theory of Mind*. Cambridge: Cambridge University Press.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press.
- Chakravartty, A. (2001). "The Semantic or Model-Theoretic View of Theories and Scientific Realism." *Synthese* 127: 325-345.
- Churchland, P. M. (1981). "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy* 78: 67-90.
- Crane, T. (2003). *The Mechanical Mind* (2nd edition). London: Routledge.
- Davies, M., and T. Stone (eds.) (1995). *Folk Psychology: The Theory of Mind Debate*. Oxford: Blackwell.
- Davies, M., and T. Stone (2001). "Mental Simulation, Tacit Theory, and the Threat of Collapse." *Philosophical Topics* 29: 127-73.
- Dennett, D. C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*. MA: MIT Press.
- Dennett, D. C. (1981). "Three Kinds of Intentional Psychology." Reprinted in *The Intentional Stance* (Cambridge MA: MIT Press, 1987), pp. 43-68.
- Fodor, J. A. (1981). *Representations*. Cambridge, MA: MIT Press.
- Fodor, J.A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.
- Gentner, D. (2002). "Mental Models, Psychology of." In N. J. Smelser and P. B. Bates (eds.), *International Encyclopedia of the Social and Behavioral Sciences*, pp. 9683-9687. Amsterdam: Elsevier.

- Giere, R. (1988). *Explaining Science: A Cognitive Approach*. Chicago: Chicago University Press.
- Giere, R. (1999). "Using Models to Represent Reality." In L. Magnani, N. J. Nersessian, and P. Thagard (eds.), *Model-Based Reasoning in Scientific Discovery*, pp. 41-57. New York: Kluwer.
- Godfrey-Smith, P. (2004). "On Folk Psychology and Mental Representation." In H. Clapin, P. Staines, and P. Slezak (eds.), *Representation in Mind: New Approaches to Mental Representation*, pp. 147-162. Amsterdam: Elsevier Publishers.
- Godfrey-Smith, P. (forthcoming[a]). "The Strategy of Model-Based Science." To appear in *Biology and Philosophy*.
- Godfrey-Smith, P. (forthcoming[b]). "Model-based Science and the Representational Theory of Mind."
- Gopnik, A., and A. Meltzoff (1997). *Words, Thoughts and Theories*. Cambridge MA: MIT Press.
- Gordon, R. M. (1995). "The Simulation Theory: Objections and Misconceptions." In Davies and Stone (1995), pp. 100-122.
- Grice, P. (1974-75). "Method in Philosophical Psychology (From the Banal to the Bizarre)." *Proceedings and Addresses of the American Philosophical Association* 48: 23-53.
- Heal, J. (1986). "Replication and Functionalism." In J. Butterfield (ed.), *Language, Mind, and Logic*. Cambridge: Cambridge University Press. Reprinted in Davies and Stone (1995), pp. 45-59.
- Hesse, M. (1966). *Models and Analogies in Science*. Notre Dame: Notre Dame University Press.
- Jackson, F., and P. Pettit (1990). "In Defence of Folk Psychology." *Philosophical Studies* 59: 31-54.
- Johnson-Laird, P. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge MA: Harvard University Press.
- Lewis, D. K. (1994). "Reduction of Mind." In S. Guttenplan (ed.), *A Companion to Philosophy of Mind*. Reprinted in D. Lewis, *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press, 1999), pp. 291-324.
- Lloyd, E. A. (1988). *The Structure and Confirmation of Evolutionary Theory*. Boulder: Greenwood Press.
- Maibom, H. (2003). "The Mindreader and the Scientist." *Mind and Language* 18: 296-315.
- Menzies, P. (forthcoming). "Reasons and Causes Revisited." Paper presented at NAMICONA workshop, Aarhus, Denmark, 2005.
- Nersessian, N. (1999). "Model-Based Reasoning in Conceptual Change." In L. Magani, N. Nersessian, and P. Thagard (eds.), *Model-Based Reasoning in Scientific Discovery*, pp. 5-22. New York: Kluwer/Plenum.
- Nichols, S., and S. P. Stich (2003). *Mindreading*. New York: Oxford University Press.
- Nisbett, R. E. (2003). *The Geography Of Thought*. New York: The Free Press.
- Ramsey, W., S. Stich, and J. Garron (1990). "Connectionism, Eliminativism and the Future of Folk Psychology." In W. Ramsey, S. Stich, and D. Rumelhart (eds.), *Philosophy and Connectionist Theory*, pp. 199-228. Hillsdale, NJ: Lawrence Erlbaum.
- Rice, S. H. (2004). *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sunderland: Sinauer.
- Roughgarden, J. (1979). *Theory of Population Genetics and Evolutionary Ecology: An Introduction*. New York: Macmillan.
- Saxe, R. (2005). "Against simulation: the argument from error." *Trends in Cognitive Sciences* 9: 174-179.
- Sellars, W. (1956/1997). *Empiricism and the Philosophy of Mind*. Cambridge MA: Harvard University Press.
- Sterelny, K. (1999). "Situated Agency and the Descent of De-

- sire." In V. Hardcastle (ed.), *Where Biology Meets Psychology: Philosophical Essays*, pp. 203-219. Cambridge, MA: MIT Press.
- Sterelny, K. (2003). *Thought in A Hostile World*. Oxford: Blackwell.
- Stich, S. P. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge MA: MIT Press.
- Stone, T., and M. Davies (eds.) (1996). *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell.
- Suppe, F. (1977). "The Search for Philosophical Understanding of Scientific Theories." In F. Suppe (ed.), *The Structure of Scientific Theories*, 2nd edition, pp. 3-232. Urbana: University of Illinois Press.
- Suppes, P. (1960). "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences." *Synthese* 12: 287-301.
- Van Fraassen, B. (1980). *The Scientific Image*. Oxford: Clarendon.
- Weisberg, M. (forthcoming). "Who is a Modeler?"
- Wimsatt, W. C. (1987). "False Models as a Means to Truer Theories." In M. Nitecki and A. Hoffmann (eds.), *Neutral Models in Biology*, pp. 23-55. Oxford: Oxford University Press.