Putting the Philosophy of Science into Mind: Knowing Minds By Models

Chuck Stieg

University of Minnesota

**Abstract**

The philosophy of science can provide fruitful contributions to other areas of philosophy. In this paper, I argue that the application of work on the nature of theories helps to resolve a long-standing dispute on the philosophy of mind over mindreading. The Theory Theory and the Simulation Theory are two competing accounts of how it is that we explain and predict the actions and mental states of others. I discuss each view as well as some of their weaknesses. I suggest that the difficulties each faces depend in part on the notion of theory supposed to be at issue. After introducing an alternative notion of theory, a model-based view, I try to show that the problems of both views are diminished and that a synthesis results.

**1. Introduction**

The ability to attribute mental states is sometimes referred to as ‘mindreading’ (see e.g. Nichols and Stich 2003) and has generated interest from various fields including philosophy, neuroscience, developmental psychology, clinical psychology and ethology. There is much debate, however, over how it is that we actually do attribute mental states such as beliefs, desires and emotions to one another. Two main rival accounts, the Theory Theory and the Simulation Theory, have been offered as explanations of mindreading and have been taken by most involved in the debate to offer quite different solutions to the question of how we attribute mental states. Both views face serious difficulties, though, and seem to lack clear articulation. I suggest that some of this lack of clarity is a result of an out of date view of theories. By adopting a more current view of scientific theories from the philosophy of science, a model-based account of theories, I hope to show that there is a middle ground between the competing views of Theory Theory and Simulation Theory. The incorporation of a model-based view provides something of a synthesis between Theory Theory and Simulation Theory and also explains away the weaknesses of each view. Moreover, scientific models have important parallels with the notion of mental models in cognitive science. The viability of mental models provides additional reasons to adopt the model view of theory in the mindreading debate.

I begin with an introduction to the Theory Theory and the Simulation Theory and problems each view faces. Then I present the models account of scientific theories and apply this account to the mindreading debate, showing how it can overcome some of the difficulties in each view. I conclude with a few additional motivations for applying the account to the mindreading debate, notably its coherence with the notion of mental models in cognitive science. While the dispute between the Simulation Theory and the Theory Theory is prima facie an empirical one, I do not here address the empirical evidence one way or the other. The aim of this paper is to examine part of the conceptual structure of the two positions. While Maibom (2003) and Godfrey-Smith (2005) also put forward the idea of folk psychology as a model, this paper differs in that it focuses on the effects of this idea on the competing views of Theory Theory and Simulation Theory. It also draws parallels between this approach and the mental models approach in cognitive science.

**2. The Theory Theory**

The account known as the Theory Theory maintains that we attribute mental states to others by applying a psychological theory to the case at hand. This psychological theory is what many philosophers and cognitive scientists refer to as “folk psychology”, the common set of assumptions and understanding people seem to have about mental states. The general interpretation of folk psychology at use in Theory Theory is as a theory that posits mental states such as beliefs and desires which are used both to explain and predict behavior. The theory is often given a functionalist rendering (see e.g. Block 1980; Carruthers & Smith 1996; Churchland 1981; Fodor 1987; Gopnik and Wellman 1995) by which it is meant that mental concepts are defined in terms of their causal relations with other mental states, input and behavior. In addition to positing mental states, the theory contains a set of principles or laws that describe the relations between mental states and environmental input and bodily output or behavior. These principles are generalizations accompanied by ceteris paribus clauses, indicating that they hold in most but not all cases. Typically, the generalizations pay special attention to the content of the mental states and so preserve the importance of the semantics of those states. Furthermore, the generalizations contain variables which allow for instantiations in particular cases. This creates at least four separate activities in the act of mindreading: 1) the possession and selection of the appropriate generalization; and 2) the instantiation of particular subjects for the variables in the generalization; 3) the filling in of the relevant details of a particular situation; and 4) the derivation of the result of the generalization, its instantiation and the relevant premises/details.

The following are examples of central principles for a theoretical belief-desire psychology:

1. *Action Principle*: An agent will act in such a way as to satisfy, or at least increase the likelihood of satisfaction, of his/her current strongest desire in the light of his/her beliefs (Botterill 1996).

2. *Perception Principle*: When an agent A attends to a situation S in a given way, and p is a fact about S that is perceptually salient in that way, then A acquires the belief that p (Botterill 1996).

3. *Satisfaction* *principle*: For all x and all p, if x hopes that p and x discovers that p obtains then x is pleased that p (Churchland 1981).

On Botterill’s (1996) view the Action Principle is the central principle of folk psychology and, he argues, should be thought of as a core principle of a Lakatosian research programme. Since core principles can only make contact with experience in conjunction with auxiliary hypotheses, or heuristic guides, the Perception Principle acts as such a heuristic guide in conjunction with particular factual data or information (Botterill 1996, 115-16). It is important here to note that Lakatos viewed the axioms, or principles, that make up the core of a scientific program as sets of sentences.

According to Theory Theory, we use theoretical generalizations, like those above, and apply the theory when we explain and/or predict our own mental states and actions and those of others. Our knowledge of our beliefs and of others’ states is thus a result of using a psychological theory. For example, employing the third principle above, if we know that Bill Clinton hopes to remain in the limelight and discovers that he has, then we can conclude or predict that Clinton will be pleased. Similarly, we apply these generalizations to ourselves.

Notice that there may be two senses of folk psychology, what Stich and Ravenscroft call internalist and externalist senses of folk psychology (Stich and Ravenscroft 1994). The external sense focuses on folk psychology as a way of talking about mental states. The theory involved is an implicit theory that introduces and defines theoretical terms such as “belief” and “desire” by the way they function in our speech, typically as a reference to a state characterized by its relations to input, output and other internal states (see, for example, Lewis 1972). The external sense, however, does not make any strong commitments about the mechanisms via which mental attributions are made. The internal sense of folk psychology claims that mental attributions are made by the use of a theory represented in some way within the brain or mind of a subject. This view posits folk psychology as a theory that serves as a mechanism by which mental states are posited in order to explain and predict the activities of ourselves and others.

In this paper, I am primarily concerned with the version of Theory Theory that adopts an internalist sense of folk psychology. In other words, it is the question of the mechanism of mindreading, and how to conceive of that mechanism, that drives my interest in applying the notion of a model to the debate between those who conceive of the mechanism as a theory and those who conceive of it as a simulation process. This does not, however, mean that the externalist sense of folk psychology is irrelevant to this discussion since the way we flesh out the mechanism, or data structure, that instantiates our folk psychological theory must rely in many ways on how we talk about mental states. Moreover, the way in which we talk about mental states can also be informed by discoveries pertaining to the internal faculty that guides our ability to attribute minds.

**2.1 Problems for the Theory Theory**

One problem for the Theory Theory is that although we may know beliefs and desires of others by theoretical inference, it seems false that we always know our own mental states by inference of this sort. Even when I have to go through some sort of reasoning process to discover what it is that I think about something, I do not find myself applying a theory to discover my beliefs and desires. I think about the qualities of the thing or person in question and make a decision about that thing or person. Will the Trailblazers make the playoffs this year? Well, they lack cohesion, they do not shoot well from the outside and their competition is stiff, so probably not. I do not believe the Trailblazers will make the playoffs. I am not even sure I needed to consult my thoughts about the Trailblazers to know that I believe they will not make the playoffs, but even if I did, it is doubtful I am theorizing about my mental states via a psychological theory in any sense resembling that discussed in the previous section.[[1]](#footnote-2)

The same seems true of desires. When a desire is presented to me, it is of an immediate sort. It may be true that I can represent my desires or goals as my own or as the way I want the world to be, but the way in which I do so is of a first-person familiarity unlike that of a product of a theory. Of course, there may be more complicated cases where we delve deeply into what is immediately present in order to find more hidden causes of our desires where we do employ a theoretical structure (e.g. as in some Freudian analysis), but this is a rarer case, more contrived and is not of the same sort of case as those at issue. While we can theorize about our mental states, and even sometimes discover them by employing some theory, this is not the typical case. This type of theory-generated discovery is also not at odds with a simulation or process approach to mental attribution since this kind of approach does not attempt to lay claim to all instances of thinking about mental states. In any event, part of the suggestion of this paper is that these kinds of differences between Theory Theory and Simulation Theory are laid bare as misconceptions and seen to be instances of both a process and a knowledge structure when conceived of in terms of employing a model to understand mental states. This point will be further articulated below.

Another problem for the Theory Theory is that it is vague. Alvin Goldman makes this charge in relation to the use of ceteris paribus clauses. Ceteris paribus clauses are statements that serve to separate off descriptions of events from possible variables that may have some effect on what is described (thus the meaning of the term: ‘all things being equal’). For example, I might say that Kassie will teach her class at noon on Monday, all things being equal (as the result of applying the generalization that teachers teach their scheduled classes, all things being equal). Adding this qualification serves to separate the potential events that might alter Kassie’s teaching her class at noon; for example, her getting ill, the destruction of her school, or her car breaking down. Notice that there are potentially many such qualifications that could play into most generalizations. These clauses, attached to the generalizations, must be left so unspecified, since there are such a large number of them, that they reduce the type of generalization to that of an approximation instead of a law (Goldman 1995).

Consider the following question and the attempt to answer it by applying a theoretical generalization. Will my wife be angry when I come home late without having called? Suppose one theoretical generalization is something like “one becomes angry when they feel they have been slighted.” Applying this to the situation, and connecting it to the relevant details, e.g. she wants to know what I am doing, that I am safe and believes that I have understood this and agreed to call her, I might deduce the answer, “yes, she will be angry.” Because of these things, she will be upset. But having the relevant details is extremely important. Consider the possibilities. “She will be upset unless she has changed her mind about what she expects from me, or unless I have actually been kidnapped or sent to the hospital, or unless she has won the lottery this evening, or become absorbed in her own work and lost track of other things.” We have to have information, perhaps guesses, about her mental states (e.g. what she wants and believes) in order to say anything about her other mental states (e.g. her anger) and in order to generate a conclusion. And we also have to have some way of representing the myriad possibilities that could affect those mental states as well as the liklihood of those obtaining. In this case, I have to have a take on these things in order to say whether my wife will feel slighted, and thus angry. In short, there are numerous possibilities that could either serve as the cause of her mental state, or that could be viewed as interfering with the outcome that our application of the folk psychological theory has predicted. The result is that the laws leave too much to be accounted for, and mindreaders cannot make a “definitive interpretive conclusion” since they would not be able to tell when these “all things being equal” conditions are met (Goldman 1995, 79). But contrary to this result, we often do make definite interpretations. This problem suggests that mindreaders do not employ theoretical generalizations or laws akin to those found as part of a theoretical structure.

Related to these problems are concerns about the characteristics of the generalizations of the theory. These include whether the supposed laws are true or accurate, whether they leave out important distinctions in mental states such as the difference between occurent and dispositional mental states and whether it is likely that such elaborations, if made in the laws, are actually possessed and employed by everyday mindreaders. Additionally, it is not even clear what the contents of the theory are supposed to be. There are a surprisingly small number of actual candidates for laws in Folk Psychology. If it is so natural for us to use Folk Psychology, we should be able to formulate its laws quite easily, especially if it is by grasping the theory that mental concepts get their meaning.[[2]](#footnote-3)

A related worry is whether the folk theory is supposed to be known consciously or unconsciously. Obviously folk psychology is not known consciously. Not many of us are aware of ever using such a theory, and children, though they seem to understand the meaning of many mental concepts, surely do not understand the role those concepts play in a theory of mind.

It is often replied that the folk theory could be known and used tacitly (see e.g. Stich and Nichols 1995). A tacit psychological theory could operate behind the scenes much as grammar principles guide our grammaticality judgments, sentence structuring, and comprehension.

This suggestion faces difficulties, though. As Botterill notes, the principles of folk psychology are unlikely to be strongly tacit in the way that the algorithms utilized in visual processing are tacit since we can often recover folk psychological principles but cannot consciously access the strongly tacit principles involved in visual processing (Botterill 1996, 113). Similarly, though native speakers may recognize certain utterances as grammatical and others as ungrammatical because they tacitly employ principles of grammar, the principles do not generally play a role in the *explanatory* practices of speakers to explain why some utterances seem grammatical and others do not. These principles are not part of the explanatory repertoire. It is different, however, with the supposed principles of folk psychology. These principles do play a role in explaining or predicting why someone did what they did. People refer to what it is the person desires and believes in that explanation. Botterill suggests that by referring to these psychological states, the target’s action is made clearer and seems explanatory because, in part, we have some awareness of the principles involved (Botterill 1996, 114). Citing them, thus, is satisfying.

If these criticisms are on the right track, then the folk psychological theory cannot be too strongly tacit since it needs to be somewhat available to awareness, and it cannot be too weakly tacit since that would suggest that more of it is available to awareness than there actually seems to be. So just what level of tacit knowledge is appropriate? Does it even make sense to speak of an unconscious theory?

I will return to these difficulties in a moment in an attempt to diagnose why these problems for the Theory Theory might arise. I will suggest that it relies on a dated view of what a theory is and that a different conception of theories, a semantic or model-based view, might alleviate some of the troubles that the Theory Theory faces and some of the alleged differences between it and the Simulation Theory. First, however, I need to describe the Simulation Theory.

**3. Simulation Theory**

Like the Theory Theory, Simulation Theory is an account of our ability to make sense of the behavior and minds of others by identifying and attributing mental states. Whereas the Theory Theory claims that the manner in which we attribute such states is theory-driven, Simulation Theory holds that these attributions are process-driven. The general idea is that one represents and arrives at the mental processes of others by generating similar states and processes in oneself. That is, we understand the minds of others by using our own minds, or psychological processes, to simulate the minds of others.

The hypothesized process takes on different forms depending on the particular brand of simulation theory under discussion (for example, that of Goldman 1995 or Gordon 1995a), but a common suggestion is that of putting oneself in the place of, or becoming, the other. To do this, one uses imagination to construct the circumstances of the other and then one’s own mental processes to generate further mental states, decisions or stances. This is often described as running one’s decision making processes ‘off-line’. These are the same decision making processes one would use for one’s self, but they are fed pretend inputs that generate some pretend output. Since the system is ‘off-line’, no real behaviors are engaged in as a result of the process.

Two major versions of the Simulation Theory are those held by Robert Gordon and Alvin Goldman. Gordon’s version maintains that when we simulate another, we are not simulating ourselves in the place of the other, rather we simulate the other. We pretend to be the other. Like an actor becoming the character they portray, we adopt the role of the other. And we do so without a list of theoretical information or laws about those characters and without introspecting anything about oneself. Gordon describes the process this way. “To simulate (another) in his situation requires an egocentric shift, a recentering of my egocentric map on that other. He becomes in my imagination the referent of the first person pronoun ‘I’…Such recentering is the prelude to transforming myself in imagination into the other much as actors become the characters they play” (Gordon 1995a, 53-4). One of Gordon’s (1995b) examples involves trying to answer why ones hiking friend has turned and run back down the trail. To do so, one pretends to be the friend and reorients his spatial location and the perceptual view as best he can to the other’s perspective. From that perspective it is clear that a grizzly bear is approaching the target. This allows the simulator to not only understand what the target perceives but also how it is perceived. It is perceived as approaching ‘me’.

Alvin Goldman’s simulation theory is somewhat different from Gordon’s. Goldman views a simulation as using oneself as a model for another instead of becoming the other. On Goldman’s view one pretends to have the same beliefs, hopes, etc. as the target has and feeds these inputs into some cognitive mechanism (e.g. an inferential mechanism). The mechanism works as it normally would and elicits some output in the form of other mental states (Goldman 1995).

A key aspect of Goldman’s view, then, is that it involves making an inference from oneself to the other in that ones own psychological processes model the target’s processes. For example, in trying to figure out a chess opponent’s likely move, one imagines oneself in the other’s situation and decides what one would then do. This serves as a model of what the actual opponent is expected to do. Gordon’s view attempts to avoid any such inference by claiming that the simulator transforms oneself into the other and so does not need to infer from ones own case to that of the other. As he puts it, we are already there.

Simulation Theory differs from the Theory Theory in its denial that mental state attribution is accomplished by use of a theoretical body of knowledge. Instead, our own psychological processes drive the attributions and in effect act as models of other minds. The process may require shifts in indexicals, spatial perspectives and epistemic perspectives. These adjustments help the simulator to approximate the simulated person’s mental states more closely and, even if these states differ from the simulator’s, to predict actions or mental states that the simulated person is likely to produce.

**3.1 Problems with the Simulation Theory**

One potential objection to the Simulation Theory is that the simulator’s processes could be tacitly operating according to a theory or some general psychological principles, in which case simulation is not a real alternative to a theory-based approach. Goldman (1995) has replied that one does not need to know how the simulation works for the simulation to produce accurate mental attributions. At the personal level, we put ourselves in the position we take the other to be in, and we simulate their situation. The key idea is that when we want to know about the reasoning of another person, we can use our own reasoning processes to discover this. In so doing, we do not utilize descriptions of cognitive mechanisms and how they interact, nor do we use laws of logic. There is no need for positing that there is a separate store of theoretical generalizations that can be called into service when thinking about others.[[3]](#footnote-4) It may be the case that our own thinking follows some particular patterns, but this is entirely different from supposing that our thinking processes are mediated by an internal theoretical structure (recall the internalist sense of folk psychology discussed earlier). While this may be so, it still seems possible that the actual procedure may operate by utilizing some tacit theory. This could be the case if, for example, the subpersonal mechanisms that underlie the simulation process are in part made up of something like a psychological theory. Part of the problem depends on the psychological level at which the mindreading process is supposed to be taking place. As with the Theory Theory, it is not always clear if the simulation process is supposed to be a process at the personal, conscious level, or whether it is supposed to be a subpersonal activity. Since the cognitive sciences are not interested in only conscious, psychological processes, the possibility of a subpersonal theoretical mechanism underlying a conscious simulation does present an objection to Simulation Theory insofar as we expect there to be a deeper explanation available in terms of the underlying mechanisms and activities. While Goldman may hold that one does not need to know what the underlying processes are in a simulation, it does not follow that they do not matter.

Another potential problem is that, as we have seen, there will need to be information included with any simulation that captures the relevant differences between the simulator and the target to be simulated. It is, again, possible that this information requires some type of theory. Whether the independent means of inferring the psychological states of others from one’s own states involves a theory, remains an open question. And if not, how it is accomplished remains an unanswered question.

A final problem for the Simulation Theory is explaining how it is that one simulates beliefs or states of those very different from the simulator.[[4]](#footnote-5) Simulating the mental states of children, animals, or abnormal adults should be very difficult for a normal adult human. Even if appropriate adjustments are made, it is unlikely that a normal cognitive mechanism will operate like that of an abnormal target. Nonetheless, we find ourselves attributing beliefs and desires to abnormal creatures that seem fairly accurate. Note that use of a general psychological theory should be more successful at this task as long as the theory contains principles about the behavior of the target under consideration.

These problems, I suggest, stem from thinking of the Simulation Theory as a foil for the Theory Theory, where the theory involved is assumed to be an elaborate network of sentence-like laws governing the psychological domain. By replacing this assumed notion of theory with the model-based view of theories, perhaps the two positions can coexist a bit more easily. In the next section I will set out this alternative view of theories and then use it to analyze the Theory Theory and the Simulation Theory.

**4. Theories as Models**

The philosophy of logical empiricism embraced the syntactic view of scientific theories, on which such a theory is a set of axioms, a set of theorems that are logical consequences of the axioms, and correspondence rules that link terms in the axioms with terms that are observationally obtained or defined. Thus conceived, a theory is comprised of a set of statements. With the demise of logical empiricism, this view of scientific theories came under attack and was criticized as being more of an ideal reconstruction of scientific theories than of an account of how theories are actually formulated, maintained, and used in science.

A prominent account of theories that has largely replaced the syntactic view is known as the semantic view of theories and also as the model-theoretic view. Ron Giere (1988), Frederick Suppe (1989), Patrick Suppes (1967) and Bas van Fraassen (1980) are important proponents and developers of this view of theories. As one might suppose, there are varieties of the semantic view. Generally speaking, this view regards theories as abstract models of relations among variables. When the parameters of the model are fixed, it becomes a specific model. A theory is a collection of models related by a family resemblance to one another. Such models can range from the abstract to the concrete in the degree to which they approximate some actual real world system. I will focus on Giere’s (1988) account to illustrate some of the main ideas in the models view of scientific theories.

Giere (1988) holds that a theory consists of (a) a group of theoretical models and (b) a hypothesis that some real system is similar to one of the theoretical models. The models are non-linguistic, abstract entities that are interpretations of equations that constitute theoretical definitions. Giere often uses examples from classical mechanics to illustrate the concepts involved in his conception. An idealized pendulum defined by Newton’s laws and the law of the pendulum is a model of some real world pendulum such as a clock. Newton’s laws of motion and law of universal gravitation serve as theoretical definitions. They are not, strictly speaking, empirical claims and require specification of concepts such as force functions. One cannot precisely apply the law of the pendulum to an actual pendulum such as a clock, for the motion of the pendulum will generally not be exactly the same as that specified by the laws. Instead, scientists make approximations of pendular motions by ignoring a variety of factors, such as other forces acting on the pendulum, thus developing an idealized description of the period of a pendulum. This idealization is a model of the equations in question and, it is important to note, only one of many possible models of the theory. What Giere terms a theoretical hypothesis claims that this model of the period of a pendulum is similar to some real system in the world.

On Giere’s view it is the models that are used to represent the world, not the equations, and it is the models that are the most important in characterizing a theory. The linguistic descriptions of those models are secondary (Giere 1989, 79-80). His view is that focusing on the definitions puts too much emphasis on linguistic entities and that this is not what scientists actually do. Furthermore, focusing on the models does not eliminate the importance of the definitions since they remain tacitly attached to the models comprising the theory. To be sure, one may labor to develop a system of equations that serve as axioms, such as those commonly offered as the axioms of Newtonian mechanics.[[5]](#footnote-6) However, Giere argues that this is to be avoided as a characterization of theories generally since other theories are unlikely to be so structured and tightly knit that the family of models can be characterized in an axiomatic manner (Giere 1988, 88). Giere also argues that axiomatic representations of classical mechanics are inaccurate portrayals of how the theory is understood and used and that it does not represent the ways physicists solve real world problems.

For our purposes the key elements of the models view of theories are as follows. First, a theory does not need to be characterized by linguistic entities. Second, the models do not have to embody strict laws to be applicable to some real world systems. They have their restricted domains built into them and some will apply more or less to some specific aspects of the world and not to others. Third, relations between the world and models need not be linguistically represented. The models are used as a way of understanding some actual system, and so are taken to resemble a part of the world in some way, but the extent to which the model resembles the real world system is left vague. The model is a particular structure, and the world, presumably, is as well. The assumption regarding the extent of fit between model and world is an additional assumption. It is entirely possible that a scientist interested in some model actually does not make any type of claim about how it applies to a real world system and instead focuses on studying the model in itself. This may be the case, for example, with some modelers studying artificial neural networks or with artificial-life models. An additional component of the models view is that scientific proficiency often amounts to competence with models and facility constructing and manipulating hypothetical structures given a certain type of schematic understanding, or common way of situating problems. The physicist, for example, may think of problems in physics as ones that involve vectors and force functions. She will have the honed skill of being able to know how to apply a general type of model to many different kinds of systems and how to describe the behavior of these systems. Familiarity with models allows for better results in applying the models and often can amount to being able to easily move from general description to particular articulation.

Characterizing scientific activity in terms of model construction and elaboration can extend beyond describing the theoretic, scientific representations as families of models. Investigatory and exploratory practices in the sciences also rely heavily on the use of models and simulation. While I discuss models in terms of representing theories in this section, I think many of the points made about the characteristics of models applies to the experimental side of modelling as well. Models in this latter sense also rely on abstraction and idealization and serve scientists to understand target systems. Indeed, there is a strong sense in which this kind of modeling is really just another form of theorizing. I want, however, to avoid any position on the relation between experimental/investigative practices and theoretical/representational activity. Since the point of this paper is to articulate the notion of theory in folk psychology, the theory-sense of model is what I am particularly interested in.[[6]](#footnote-7)

**5. Applying the Model View**

Having noted some of the characteristics of the models view of theories, I would like to see what the result would be if we were to think of folk psychology in the Theory Theory as a theory in Giere’s sense. It seems that the assumption that theories are composed of axioms and theorems is crucially involved in the debate between Theory theorists and Simulation theorists. For the folk psychological theory posited by Theory theorists is apparently composed of sets of universally quantified sentences (axioms) that require descriptions of observed facts (initial conditions) in order to derive particular testable predictions and explanations. This conception is startlingly similar to that of the Logical Empiricist view of theories and explanation.

Viewing theories as sets of statements creates some of the problems Theory Theory faces. For it makes every instance of knowing something about ourselves or others a matter of inference, a movement from one set of statements to another, universally quantified statements governing our thoughts. What are these statements? In what language are they written? And how did they get there? We are unable to say. Furthermore, the attempt to avoid such problems inclines simulationists to deny that theory plays an important role in mindreading. But this move is unwarranted because it is possible that the sort of theory at work in folk psychological practices is a non-linguistic family of models. A model-based folk psychology could satisfy both the need for a theoretical component, since theories are often models, and the need for a process-oriented explanation of folk psychological practices, since models are dynamic structures used for simulating and representing phenomena.

The idea being suggested is that folk psychological practices might be essentially model-based. Folk psychology is a family of models that capture the relations between social/environmental conditions, common mental states such as beliefs and desires and particular behaviors. The models making up the family are related by resemblance to one another in the common function of behavioral and psychological interpretation, whether it be explanatory, predictive or simply one that produces a sense of understanding for the user.

Models may vary by the relations they represent and the complexity of those relations. Models of canine behavior and human behavior, for example, will differ in some ways and resemble one another in other ways. For instance, models of canine behavior may represent relations of dominance as playing important roles in the causation of states of fear and aggression while human folk psychological models do not. Models brought to bear on unfamiliar humans from different cultures might be closer to ones model of familiar cultures but still differ in certain ways. The act of interpreting others’ minds is accomplished by taking the general model, or most familiar model, of folk psychology, and applying it to the target at hand. Like the scientist experienced with use of a model, the mindreader can be adept at applying a model to its target system as it develops an understanding of the target’s mental states. This resembles the use of models in that the folk psychological model resembles a rough schema of a psychological system. The general schema can be readily adopted to fit the nuances of particular circumstances just as the game theory modeller can alter the type of model they use to understand the interactions of individuals within a population. The speed with which normal mindreaders attribute complex personality traits and general character profiles to others might be best explained by this application of a basic folk psychological model to varying scenarios.

If the psychological principles we use as part of a general theory are sentence-like, they would likely be large in number and would cause computations dependent on them to be lengthy and cumbersome. They would also require a strategy by which we could activate the appropriate laws in appropriate circumstances. By getting rid of laws in this sense, we can move towards making better sense of our actual cognitive mechanisms. The notions of cognitive schemas and maps are common in cognitive science and have been explored in a variety of animals. If, as Nichols and Stich (2003) argue, our ability to mindread has its roots in mechanisms that served rudimentally similar purposes earlier in our evolutionary history, then it makes sense that there should be some continuity in the structure of those mechanisms. Moreover, if it is true that scientific theories are actually best thought of as representational devices illustrated best as families of models, we can also bring into union the way cognitive agents represent and the representations we create.

**5.1 Mental Models, Folk Psychological Models and Scientific Models**

Before examining some of the effects of the model view of folk psychology on the simulation-theory debate, I want to take a brief look at some of the general themes of the mental models literature in cognitive science. Since the internal sense of folk psychology implicates an internally represented mechanism through which people interpret minds, the mental model literature serves as a plausible source of ideas regarding the nature of such mechanisms.

Mental models have been proposed as ways of explaining how people perform a variety of tasks including, but not limited to, decision making, perceptual discrimination and navigation, reasoning, and knowledge of causal relationships. Two of the most well-known proponents of mental models are Johnson-Laird (e.g. Johnson-Laird 1983) and Gentner and Stevens (e.g. Gentner and Stevens 1983). Gentner and Stevens’ use of mental models focuses on understanding how physical systems work, and the models serve as knowledge representations of the target system. Johnson-Laird’s research centers around the use of mental models in the process of solving deductive reasoning problems and situates itself as an alternative to theories that posit the solution of deductive problems by the use of logical inference rules to propositions. On his view, mental models are like diagrams that constitute particular possible combinations of premises and conclusions and work to provide counterexamples of potential deductive arguments to the reasoner. Johnson-Laird argues that mental models are the key representational figures that comprise the basic structure of cognition and are used in a variety of contexts, including the representation of social and psychological actions (Johnson-Laird 1983, 397).

The key features of mental models that parallel the notion of folk psychology as a model are the following. Mental models are internal representations of some other state of affairs or phenomenon. As such, they embody particular structural properties of the target system, be they spatial, temporal, causal or functional properties. Similarly, folk psychology is a representation of the relations of mental states, personalities, social/environmental features and behavior. Mental models are schematic. They contain selective information about the target system and do not represent all aspects of those systems. Likewise, folk psychology represents only certain aspects of organisms, i.e. their psychology and behavior, and does not, typically, attempt complete treatments of its psychological targets. Mental models, like folk psychological practices, may also contain errors and lead to failures in tasks; for example, in judgments of the validity of an argument. As schema, mental models are not only incomplete, but they also are unstable insofar as they are able to change and evolve over time. Mental models for novel situations can be constructed quickly from other schema and past experience, and can be used to provide simplified explanation and prediction of complex phenomena. This use of mental models is a key element of the utility of folk psychological interpretation, i.e. the ease with which it provides a grasp on the complex behavior of agents.

The above elements of mental models and folk psychology are also aspects of scientific models. Scientific models also are incomplete schematic representations of some target system. They function as idealizations that focus on particular aspects of complex systems in order to provide the scientist with a manageable approach to understand, explain and predict the behavior of complex systems. The models can be put into the service of understanding new phenomena and can be altered to instantiate different hypothetical, actual and novel circumstances. The level of skill, or facility, in use of a model can vary from scientist to scientist just as it can from mindreader to mindreader. Furthermore, scientific models are subject to error and change, and are constantly evolving, just as mental models may be adjusted with experience and mindreaders may alter their models of others’ mental capacities.[[7]](#footnote-8)

**5.2 Consequences for the Mindreading Debate**

If the theory of folk psychology is conceived of as a family of models in Giere’s sense, many of the difficulties of both Theory Theory and Simulation Theory are eased.[[8]](#footnote-9) Adopting the model-based view rids the reliance of theory on linguistic ties. For the Theory Theory, this diffuses some of the difficulty in explaining why we do not always know our own mental states *via* theoretical inference even though we supposedly should employ this method for both ourselves and others. The key here is that we do not use an inference guided by axioms to discover our own or another’s mental states. Since models are nonlinguistic entities, they do not require inferences for the models to represent real world systems. Perhaps when we engage in mindreading, we employ a model of some set of circumstances and situate the actors, whether ourselves or others, in that model such that no laws or generalizations are needed to infer what the outcome will be in that model.

The vagueness of the theory is also downplayed since the ceteris paribus clauses that are imagined to be necessarily added to each generalization are built into the variety of models making up the theory. Recall from section 2.1 the problem Theory Theory faces in accounting for the “all things being equal” qualifications of its generalizations. Given the multitude of potential variables that can affect the reasoning and mental states of subjects, any law that ranges over human psychological states will need to be able to protect against the perturbations such variables might have. A real world agent engaged in mindreading will need to be able to attach the qualifications of the target’s situation to their use of the generalization in order to come to an accurate prediction or a reasonable explanation of the target’s mental states. This list of ceteris paribus clauses would likely be extremely lengthy and unwieldy. If the way we reason about others is based on the use of models, though, a solution to this problem presents itself. The clauses are part of the model, providing its parameters and assumptions about the system. Thus, it will exclude certain things as part of its structure and alleviate the need for laws to list the exceptions. Just as the idealized model of a pendulum excludes other forces from its description of the pendulum (and thus avoids explicitly describing all of the excluded forces), a psychological model brought to bear in a particular circumstance can provide a characterization of the target individual without having to consider all potential, intervening factors. For example, in interpreting the behavior of a person who is trembling uncontrollably, we might conclude that they think they are in danger and extremely afraid of something, ignoring the possibility that they are very cold or even suffering from a neurological condition.[[9]](#footnote-10)

A family of models also helps address the accuracy of the laws. Insofar as there are any laws, they are embodied in the model. In that sense, they are true, or accurate, in the model even though they may not fit perfectly the thing being modeled. It may even be possible to develop, as part of our maturing cognitive economy, new models that tweak prior ones to accommodate differences in contexts of the real world.

The difficulty in fully stating the laws of the theory may be avoided by noting that either there are no laws in the traditional folk psychological sense, in which case not being able to set them out is no problem, or the laws are merely descriptions of different models. In the latter case, it may not be surprising to find that we cannot fully articulate those principles, but rather we have a sense of familiarity with using the models. Perhaps articulating the appropriate descriptions of our cognitive models is a research agenda of cognitive science, but adopting the model-based view of theories makes it clear that we do not need to possess such descriptions in order to use the models. Similarly, the notion of unconscious models may strike some as less mysterious than unconscious sentences.

As for the Simulation Theory, a models view allows simulationists to accept that an underlying theory might be responsible for running simulations. It could be a theory, but the theory is a model, or group of models, and simulates activity within that model. The use of models goes some way towards preserving the general, key idea that mindreading is process-driven, i.e. an activity based on a general set of reasoning mechanisms, here, the folk psychological models. In the theory-simulation debate, this is supposed to be in contrast to performing interpretation by the application of theoretical generalizations.

Using models as the basis of the theory involved in mindreading also may help to account for beliefs of those very different from ourselves since theories are made up of families of models that may differ in ways relevant to interpreting various types of beings. Traditional simulation accounts of mindreading have difficulty explaining how a person can know anything about radically different minded individuals, e.g. someone with autism or another species, since “putting oneself in the other’s place” would not adequately capture the other’s mental states given the supposed differences between their psychological economies. Theory approaches are better able to deal with this problem since they can allow for different theoretical generalizations for different domains. The model-theoretic view utilizes this basic insight of the theory approach and may even improve upon it by providing for a possible mechanism by which the appropriate models are activated for a given circumstance: the relation of mental model schema to perceptual scenarios and the experienced facility of model users with relevant model instantiation. It should not be ignored, either, that there is a limit to our ability to interpret alien beings. We do not perfectly explain and predict each other, let alone radically different creatures. Our models may accommodate some interpretation, but there is no reason to suppose that we should be able to successfully interpret every being.

Though the view I have suggested here may seem like Simulation Theory in disguise, the general view that results from conceiving of psychological theory as model-based differs importantly from Simulation Theory. Whereas Simulation Theory is often characterized as information-poor, the model-based view is information-rich. It allows for important informational inputs into the psychological attribution process. Furthermore, it allows that much of this information could involve theoretical assumptions. Also, the model-based view makes space for characterizing theoretical generalizations, here called theoretical definitions, embodied in the structure of models. Simulation Theory, on the other hand, often leaves unexpressed the mechanisms of, or what is taking place during, psychological attribution.

By adopting the view that theories are families of models, it is possible to close the gap between Simulation Theory and Theory Theory. They do not have to remain polar opposites in competition for the explanation of our mindreading abilities. While it is possible that we utilize both strategies as they are traditionally described, we can bring the two views into one basic strategy by adopting the models view. Using a theory, on this view, is using a model or family of models. Using a family of models to represent targets, their actions and their mental states involves situating them in something like an idealized picture. Situating them in this way is a form of running a process and not just manipulating generalizations. Thus, mindreading via models can be both theory-driven and process-driven at the same time.

**References**

Block, N. ed. 1980: *Readings in Philosophy of Psychology*, vol. 1. Cambridge, MA:

Harvard University Press.

Botterill, G. 1996: Folk psychology and theoretical status. In P. Carruthers and P. Smith (eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Carruthers, P. and Smith, P. (eds.) 1996: *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Churchland, P. 1981: Eliminative materialism and propositional attitudes. *The Journal of Philosophy*, 78, 67-90.

Davies, M. and Stone, T. (eds.) 1995: *Folk Psychology*. Cambridge, MA: Blackwell Publishers, Inc.

Davies, M. & Stone, T., (eds.) 1995: *Mental Simulation*. Oxford: Blackwell Publishers, Inc.

Fodor, J. 1987: *Psychosemantics*. Cambridge, MA.: MIT Press.

Gentner, D. and Stevens, A. 1983. *Mental Models*. New Jersey: Erlbaum.

Giere, R. 1988: *Explaining Science: a Cognitive Approach*. Chicago: University of Chicago Press.

Godfrey-Smith, P. 2005: Folk psychology as a model. *Philosophers’ Imprint*, 5, 1-16.

Goldman, A. 1995: Interpretation Psychologized. In M. Davies and T. Stone (eds.), *Folk Psychology*. Cambridge, MA: Blackwell Publishers, Inc.

Gopnik, A. and Wellman, H. 1995: Why the child’s theory of mind really is a theory. In M. Davies and T. Stone (eds.), *Folk Psychology*. Cambridge, MA: Blackwell Publishers.

Gordon, R. 1995a: Simulation without introspection or inference from me to you. In M. Davies and T. Stone (eds.), *Mental Simulation*. Oxford: Blackwell Publishers.

Gordon, R. 1995b: The simulation theory: objections and misconceptions. in

*Folk Psychology*. Davies & Stone (eds.) Cambridge, MA: Blackwell Publishers.

Greiner, W. and Bromley, S.A. 2003: *Classical Mechanics: Point Particles and Relativity*. New York: Springer.

Johnson-Laird, P. 1983: *Mental Models*. Cambridge, MA: Harvard University Press.

Lewis, D. 1972: Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50, 249-58.

Maibom, H. 2003: The mindreader and the scientist. *Mind and Language*, 18, 296-315.

Nersessian, N.J. 2002: The cognitive basis of model-based reasoning in science. In P. Carruthers, S. Stich and M. Siegal (eds.), *The Cognitive Basis of Science*.

Nichols, S. and Stich, S. 2003: *Mindreading: an Integrated Account of Pretence, Self-awareness, and Understanding Other Minds*. Oxford: Clarendon Press.

Segal, G. 1994: Belief: epistemology of. In S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*. Oxford: Blackwell Publishers.

Stich, S. and Nichols, S. 1995: Folk psychology: simulation or tacit theory? In M. Davies and T. Stone (eds.), *Folk Psychology*. Cambridge, MA: Blackwell Publishers.

Stich, S. and Ravenscroft, I. 1994: What is folk psychology? *Cognition*, 50, 447-68.

Suppe, F. 1989: *The Semantic Conception of Theories and Scientific Realism*. Urbana & Chicago: University of Illinois Press.

Suppes, P. 1967: What is a scientific theory? in S. Morgenbesser (ed.), *Philosophy of* *Science Today*. New York: Basic Books.

Van Fraassen, B. 1980: *The Scientific Image*. Oxford: Clarendon Press.

1. Segal (1994) makes the same point. [↑](#footnote-ref-2)
2. See Maibom (2003) for an argument that uses a similar point to argue that folk psychology does not consist of universal generalizations. [↑](#footnote-ref-3)
3. This is sometimes called the Parsimony Argument for Simulation Theory. [↑](#footnote-ref-4)
4. Again, Segal (1994) makes this point as well. [↑](#footnote-ref-5)
5. See, for example, Greiner and Bromley (2003, chapter 16) for a detailed discussion of Newton’s three axioms: the law of inertia, the fundamental equation of dynamics, and the interaction law. [↑](#footnote-ref-6)
6. I am also not interested in making the strong claim that all of theorizing in science is model-based. This is not necessary in order to think of folk psychological theory as model-based. [↑](#footnote-ref-7)
7. See Nersessian (2002) for an argument that model-based reasoning in science has its roots in cognitive model manipulation which itself is intimately connected to mechanisms used in perceptual-motor processing. [↑](#footnote-ref-8)
8. Most of the difficulties I claim are overcome by the application of the model-based view of theories are somewhat distant from the empirical challenges both the Theory Theory and Simulation Theory face. This has been intentional since one, the project here is conceptual, and two, it is unclear to what extent these views fail to account for the empirical data. Indeed, how to interpret much of the empirical data is a point of contention. For example, while Gopnik and Wellman (1995) argue that children develop through stages that involve developing from a nonrepresentational theory of mind to a fully representational one, Nichols and Stich (2003, 107-16) argue that the evidence does not show this at all. Nevertheless, there are clearly some empirical results that any theory of mindreading must be able to account for. Nichols and Stich (2003) provide a checklist of this sort, and though it is beyond the scope of this paper, I believe a model-based view of mindreading, properly developed, could account for such data. [↑](#footnote-ref-9)
9. The description provided by the model might be wrong in a particular case to ignore some factor, but it is a fact of both models and folk psychological practice that we often do quickly describe/interpret someone and may be inaccurate in our assessment. [↑](#footnote-ref-10)