This article appeared in Mind & Language, v. 7, no. 1, 1992, 35-71.

Archived at Website for the Rutgers University Research Group on Evolution and Higher Cognition.

Folk Psychology: Simulation or Tacit Theory?

Stephen Stich
Department of Philosophy and Center for Cognitive Science
Rutgers University
New Brunswick, NJ 08901
stich@ruccs.rutgers.edu

and

Shaun Nichols
Department of Philosophy
College of Charleston
Charleston, SC 29424
nichols@cofc.edu

Contents

---

1. Introduction

A central goal of contemporary cognitive science is the explanation of cognitive abilities or capacities. [Cummins 1983] During the last three decades a wide range of cognitive capacities have been subjected to careful empirical scrutiny. The adult's ability to produce and comprehend natural language sentences and the child's capacity to acquire a natural language were among the first to be explored. [Chomsky 1965, Fodor, Bever & Garrett 1974, Pinker 1989] There is also a rich literature on the ability to solve mathematical problems [Greeno 1983], the ability to recognize objects visually [Rock 1983, Gregory 1970, Marr 1982], the ability to manipulate and predict the behavior of middle sized physical objects [McClosky 1983, Hayes 1985], and a host of others.

In all of this work, the dominant explanatory strategy proceeds by positing an internally represented "knowledge structure" - typically a body of rules or principles or propositions - which serves to guide the execution of the capacity to be explained. These rules or principles or propositions are often described as the agent's "theory" of the domain in question. In some cases, the theory may be partly accessible to consciousness; the agent can tell us some of the rules or principles he is using. More often, however, the agent has no conscious access to the knowledge guiding his behavior. The theory is "tacit" [Chomsky 1965] or "sub-doxastic" [Stich 1978]. Perhaps the earliest philosophical account of this explanatory strategy is set out in Jerry Fodor's paper, "The Appeal to Tacit Knowledge in Psychological Explanations." [Fodor 1968]. Since then the idea has been elaborated by Dennett [1978a], Lycan [1981, 1988], and a host of others.

Among the many cognitive capacities that people manifest, there is one cluster that holds a particular fascination for philosophers. Included in this cluster is the ability to *describe* people and their behavior (including their linguistic behavior) *in intentional terms* - or to "interpret" them, as philosophers sometimes say. We exercise this ability when we

describe John as *believing that the mail has come*, or when we say that Anna *wants to go to the library*. By exploiting these intentional descriptions, people are able to offer *explanations* of each other's behavior (Susan left the building *because* she believed that it was on fire) and to *predict* each other's behavior, often with impressive accuracy. Since the dominant strategy for explaining any cognitive capacity is to posit an internally represented theory, it is not surprising that in this area, too, it is generally assumed that a theory is being invoked. [Churchland 1981 & 1989, Fodor 1987, Sellars 1963. See also Olson et. al. 1988] The term "folk psychology" has been widely used as a label for the largely tacit psychological theory that underlies these abilities. During the last decade or so there has been a fair amount of empirical work aimed at describing or modeling folk psychology and tracking its emergence and development in the child. [D'Andrade 1987, Leslie 1987, Astington et. al. 1988]

Recently, however, Robert Gordon, Alvin Goldman and a number of other philosophers have offered a bold challenge to the received view about the cognitive mechanisms underlying our ability to describe, predict and explain people's behavior. [Goldman 1989, Gordon 1986, Gordon unpublished a, Gordon unpublished b, Montgomery 1987, Ripstein 1987, Heal 1986][(1)] Though they differ on the details, these philosophers agree in denying that an internally represented folk psychological theory plays a central role in the exercise of these abilities. They also agree that a special sort of mental *simulation* in which we use ourselves as a model for the person we are describing or predicting, will play an important role in the correct account of the mechanisms subserving these abilities. In this paper, although we will occasionally mention the views of other advocates of simulation, our principal focus will be on Gordon and Goldman.

If these philosophers are right, two enormously important consequences will follow. First, of course, the dominant explanatory strategy in cognitive science, the strategy that appeals to internally represented knowledge structures, will be shown to be mistaken in at least one crucial corner of our mental lives. And if it is mistaken there, then perhaps theorists exploring other cognitive capacities can no longer simply take the strategy for granted.

To explain the second consequence we will need a quick review of one of the central debates in recent philosophy of mind. The issue in the debate is the very existence of the intentional mental states that are appealed to in our ordinary explanations of behavior - states like believing, desiring, thinking, hoping, and the rest. *Eliminativists* maintain that there really are no such things. Beliefs and desires are like phlogiston, caloric and witches; they are the mistaken posits of a radically false theory. The theory in question is "folk psychology" - the collection of psychological principles and generalizations which, according to eliminativists (and most of their opponents) underlies our everyday explanations of behavior. The central premise in the eliminativist's argument is that neuroscience (or connectionism or cognitive science) is on the verge of demonstrating persuasively that folk psychology is false. But if Gordon and Goldman are right, they will have pulled the rug out from under the eliminativists. For if what underlies our ordinary explanatory practice is not a theory at all, then obviously it cannot be a radically false theory. There is a certain delightful irony in the Gordon/Goldman attack on eliminativism. Indeed, one might almost view it as attempting a bit of philosophical jujitsu. The eliminativists claim that there are no such things as beliefs and desires because the folk psychology that posits them is a radically false theory. Gordon and Goldman claim that the theory which posits a tacitly known folk psychology is *itself* radically false, since there are much better ways of explaining people's abilities to interpret and predict behavior. Thus, if Gordon and Goldman are right, *there is no such thing as folk psychology*! [Gordon 1986, p. 170; Goldman 1989, p. 182]

There can be no doubt that if Gordon and Goldman are right, then the impact on both cognitive science and the philosophy of mind will be considerable. But it is a lot easier to doubt that their views about mental simulation are defensible. The remainder of this paper will be devoted to developing these doubts. Here's the game plan for the pages to follow. In Sections 2 and 3, we will try to get as clear as we can on what the simulation theorists claim. We'll begin, in Section 2, with an account of the special sort of simulation that lies at the heart of the Gordon/Goldman proposal. In that section our focus will be on the way that simulation might be used in the *prediction* of behavior. In Section 3, we'll explore the ways in which mental simulation might be used to explain the other two cognitive capacities that have been of special interest to philosophers: *explaining* behavior and producing *intentional descriptions* or *interpretations*. We'll also consider the possibility that simulation might be used in explaining the *meaning* of intentional terms like `believes,' and `desires'. Since the accounts of simulation that Gordon and Goldman have offered have been a bit sketchy, there will be a lot of filling in to do in Sections 2 and 3. But throughout both Sections, our goal will be sympathetic interpretation; we've tried hard not to build straw men. In the following two Sections, our stance turns critical. In Section 4, we will do our best to assemble all the arguments offered by Gordon and Goldman in support of their simulation theory, and to explain why none of them are convincing. In Section 5 we will offer two arguments of our own, aimed at showing why, in light of currently available evidence, the simulation theory is very implausible indeed. Section 6 is a brief conclusion.

2. Predicting Behavior: Theory, Simulation and Imagination

Suppose that you are an aeronautical engineer and that you want to predict how a newly built plane will behave at a certain speed. There are two rather different ways in which you might proceed. One way is to sit down with pencil and paper, a detailed set of specifications of the plane, and a state of the art textbook on aerodynamic theory, and try to calculate what the theory entails about the behavior of the plane. Alternatively, you could build a model of the plane, put it in a wind tunnel, and observe how it behaves. You have to use a bit of theory in this second strategy, of course, since you have to have some idea which properties of the plane you want to duplicate in your model. But there is a clear sense in which a theory is playing the central role in the first prediction and a model or simulation is playing a central role in the second.[(2)]

Much the same story could be told if what you want to do is predict the behavior of a person. Suppose, for example, you want to predict what a certain rising young political figure would do if someone in authority tells him to administer painful electric shocks to a person strapped in a chair in the next room. One approach is to gather as much data as you can about the history and personality of the politician and then consult the best theory available on the determinants of behavior under such circumstances. Another approach is to set up a Milgram-style experiment and observe how some other people behave. Naturally, it would be a good idea to find experimental subjects who are psychologically similar to the political figure whose behavior you are trying to predict. Here, as before, theory plays a central role in the first prediction, while a simulation plays a central role in the second.

In both the aeronautical case and the psychological case, we have been supposing that much of the predicting process is carried on outside the predictor. You do your calculations on a piece of paper; your simulations are done in wind tunnels or laboratories. But, of course, it will often be possible to internalize this process. The case is clearest when a theory is being used. Rather than looking in a textbook, you could memorize the theory, and rather than doing the calculations on a piece of paper, you could do them in your head. Moreover, it seems entirely possible that you could learn the theory so well that you are hardly conscious of using it or of doing any explicit calculation or reasoning. Indeed, this, near enough, is the standard story about a wide variety of cognitive capacities.

A parallel story might be told for predictions using simulations. Rather than building a model and putting it in a wind tunnel, you could *imagine* the model in the wind tunnel and see how your imaginary model behaves. Similarly, you could *imagine* putting someone in a Milgram-style laboratory and see how your imaginary subject behaves. But obviously there is a problem lurking here. For while it is certainly possible to imagine a plane in a wind tunnel, it is not at all clear how you could successfully imagine the behavior of the plane unless you had a fair amount of detailed information about the behavior of planes in situations like this one. When the simulation uses a real model plane, the world tells you how the model will behave. You just have to look and see. But when you are only imagining the simulation, there is no real model for you to look at. So it seems that you must have an internalized knowledge structure to guide your imagination. The theory or knowledge structure that you are exploiting may, of course, be a tacit one, and you may be quite unaware that you are using it. But unless we suppose your imagination is guided by some systematic body of information about the behavior of planes in situations like this one, the success of your prediction would be magic.

When you are imagining the behavior of a person, however, there are various ways in which the underlying system might work. One possibility is that imagining the behavior of a person is entirely parallel to imagining the behavior of a plane. In both cases your imagination is guided by a largely tacit theory or knowledge structure. But there is also a very different mechanism that might be used. In the plane case, you don't have a real plane to observe, so you have to rely on some stored information about planes. You do, however, have a real, human cognitive system to observe - your own. Here's a plausible, though obviously over-simplified, story about how that system normally works:

> At any given time you have a large store of beliefs and desires. Some of the beliefs are derived from perception, others from inference. Some of the desires (like the desire to get a drink) arise from systems monitoring bodily states, others (like the desire to go into the kitchen) are "sub-goals" generated by the decision making (or "practical reasoning") system. The decision making system, which takes your beliefs and desires as input, does more than generate sub-goals, it also somehow or other comes up with a decision about what to do. That decision is then passed on to the "action controllers" - the mental mechanisms responsible for sequencing and coordinating the behavior necessary to carry out the decision. Rendered boxologically, the account just sketched appears in Figure 1.

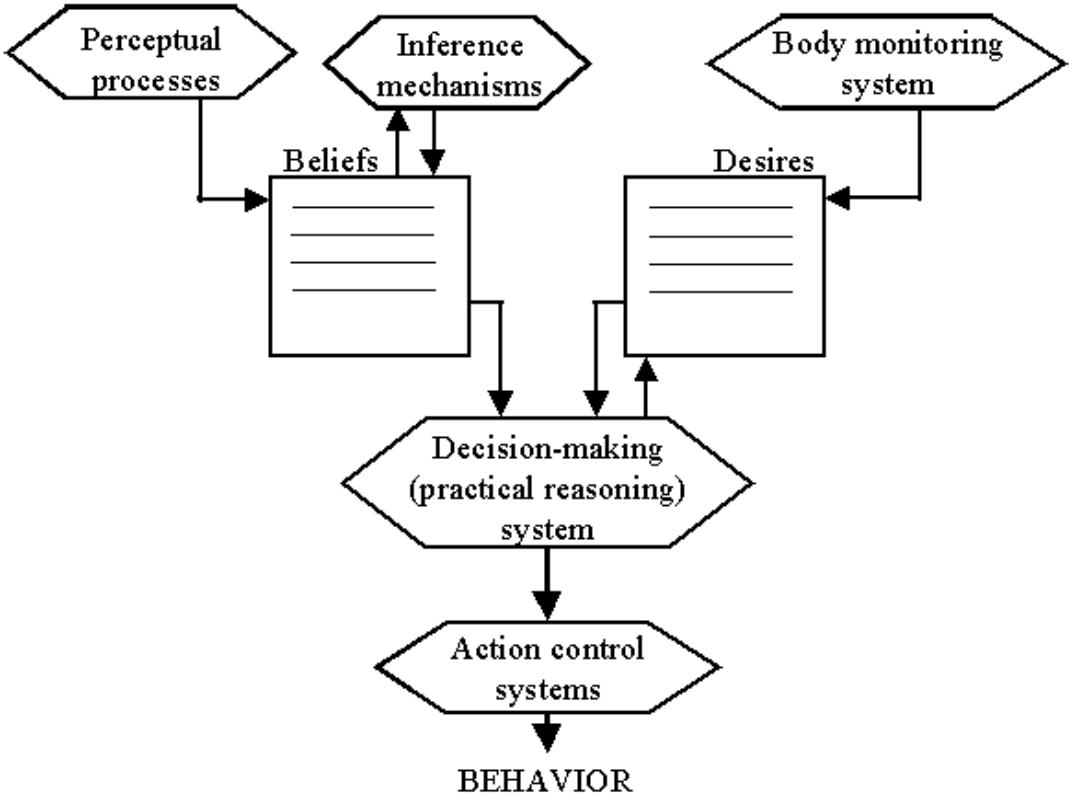

Figure 1

Now suppose that it is possible to take the decision making system "off-line" by disengaging the connection between the system and the action controllers. You might then use it to generate decisions that you are not about to act on. Suppose further that in this off-line mode, you can feed the decision making system some hypothetical or "pretend" beliefs and desires - beliefs and desires that you do not actually have, but that the person whose behavior you're trying to predict does. If all this were possible, you could then sit back and let the system generate a decision. Moreover, if your decision making system is similar to the one in the person whose behavior you're trying to predict, and if the hypothetical beliefs and desires you've fed into your system off-line are close to the ones that he has, then the decision that your system generates will often be similar to the one that his system generates. There is no need for a special internalized knowledge structure here; no tacit folk psychological theory is being used. Rather, you are using (part of) your own cognitive mechanism as a model for (part of) his. Moreover, just as in the case where the prediction exploits a theory, this whole process may be largely unconscious. It may be that all you are aware of is the prediction itself. Alternatively, if you consciously imagine what the target of your prediction will do, it could well be the case that your imagination is guided by this simulation rather than by some internally represented psychological theory.

We now have at least the outline of an account of how mental simulation might be used in predicting another person's behavior. An entirely parallel story can be told about predicting our own behavior under counterfactual circumstances. If, for example, I want to know what I would I do if I believed that there was a burglar in the basement, I can simply take my decision making system off-line and provide it with the pretend belief that there is a burglar in the basement.[(3)]

In the next section we'll try to get clear on how this process of simulation might be used in explaining various other cognitive capacities. But before attending to that task, we would do well to assemble a few quotes to confirm our claim that the story we've told is very close to the one that those we'll be criticizing have in mind. Gordon is much more explicit than Goldman on the use of simulation in prediction. Here's a passage from his 1986 paper:

> [O]ur decision-making or practical reasoning system gets partially disengaged from its `natural' inputs and fed instead with suppositions and images (or their `subpersonal' or `sub-doxastic' counterparts). Given these

> artificial pretend inputs the system then `makes up its mind' what to do. Since the system is being run off-line, as it were, disengaged also from its natural output systems, its `decision' isn't actually executed but rather ends up as an anticipation ... of the other's behavior. [Gordon 1986, p. 170]

And another, this time from an unpublished manuscript contrasting his view to Fodor's:

> The Simulation Theory as I present it holds that we explain and predict behavior not by applying a theory but simply by exercising a skill that has two components: the capacity for practical reasoning -- roughly, for making decisions on the basis of facts and values -- and the capacity to introduce "pretend" facts and values into one's decision-making typically to adjust for relevant differences in situation and past behavior. One predicts what the other will decide to do by making a decision oneself -- a "pretend" decision, of course, made only in imagination -- after making such adjustments. [Gordon unpublished, a, MS p. 3]

Gordon later suggests that the capacity to simulate in this way may be largely innate:

> [Evidence] suggests that the readiness for simulation is a prepackaged "module" called upon automatically in the perception of other human beings.(4) It suggests also that supporting and complementing the conscious, reportable procedure we call putting ourselves in the other's place, those neural systems that are responsible for the formation of emotions and intentions are, often without our knowledge, allowed to run off-line: They are partially disengaged from their "natural" inputs from perception and memory and fed artificial pretend inputs; uncoupled also from their natural output systems, they terminate not as intentions and emotions but as anticipations of, or perhaps just unconscious motor adjustments to, the other's intentions, emotions, behavior. [Gordon unpublished a, MS p. 5]

3. Other Uses For Simulation: Explanation, Interpretation and the Meaning of Intentional Terms

Let's turn, now, to people's ability to offer *intentional explanations* of other people's actions. How might mental simulation be used to account for that ability?

Consider, for example, a case similar to one proposed by Gordon.(5) We are seated at a restaurant and someone comes up to us and starts speaking to us in a foreign language. How might simulation be exploited in producing an intentional explanation for that behavior?

One proposal, endorsed by both Gordon and Goldman, begins with the fact that simulations can be used in predictions, and goes on to suggest that intentional explanations can be generated by invoking something akin to the strategy of analysis-by-synthesis. In using simulations to predict behavior, hypothetical beliefs and desires are fed into our own decision making system (being used "off-line" of course), and we predict that the agent would do what we would decide to do, given those beliefs and desires. A first step in *explaining* a behavioral episode that has already occurred is to see if we can find some hypothetical beliefs and desires which, when fed into our decision mechanism, will produce a decision to perform the behavior we want to explain.

Generally, of course, there will be *lots* of hypothetical beliefs and desires that might lead us to the behavior in question. Here are just a few:

(a) If we believe someone only speaks a certain foreign language and we want to ask him something, then we would decide to speak to him in that language.

(b) If we want to impress someone and we believe that speaking in a foreign language will impress him, then we will decide to speak to him in that language.

(c) If we believe that speaking to someone in a foreign language will make him laugh, and if we want to make him laugh, then we will decide to speak to him in that language.

And so on. Each of these simulation-based predictions provides the kernel for a possible explanation of the behavior we are trying to explain. To decide among these alternative explanations, we must determine which of the input belief/desire pairs is most plausibly attributed to the agent. Some belief/desire pairs will be easy to exclude. Perhaps the agent is a dour fellow; he never wants to make anyone laugh. If we believe this to be the case, then (c) won't be very plausible. In other

cases we can use information about the agent's perceptual situation to assess the likelihood of various beliefs. If Mary has just made a rude gesture directly in front of the agent, then it is likely the agent will believe that Mary has insulted him. If the rude gesture was made behind the agent's back, then it is not likely he will believe that she has insulted him. In still other cases, we may have some pre-existing knowledge of the agent's beliefs and desires. But, as both Goldman and Gordon note, it will often be the case that there are lots of alternative explanations that can't be excluded on the basis of evidence about the agent's circumstances or his history. In these cases, Goldman maintains, we simply assume that the agent is psychologically similar to us - we attribute beliefs that are "natural for us" [Goldman 1989, p. 178] and reject (or perhaps do not even consider) hypotheses attributing beliefs that we consider to be less natural. [Goldman 1989, pp. 178-9] Gordon tells much the same story.

> No matter how long I go on testing hypotheses, I will not have tried out *all* candidate explanations of the [agent's] behavior. Perhaps some of the unexamined candidates would have done at least as well as the one I settle for, if I settle: perhaps indefinitely many of them would have. But these would be `far fetched', I say intuitively. Therein I exhibit my inertial bias. The less `fetching' (or `stretching', as actors say) I have to do to track the other's behavior, the better. I tend to *feign* only when necessary, only when something in the other's behavior doesn't fit. This inertial bias may be thought of as a `least effort' principle: the `principle of least pretending'. It explains why, other things being equal, I will prefer the less radical departure from the `real' world - i.e. from what I myself take to be the world. [Gordon 1986, p. 164][(6)]

Though the views endorsed by Gordon and Goldman are generally very similar, the two writers do differ in their emphasis. For Gordon, prediction and explanation loom large, while for Goldman, the capacity to *interpret* people, or to describe them in intentional terms, is given pride of place. Part of the story Goldman tells about simulation-based intentional description relies on the account of simulation-based explanation that we have just sketched. One of the ways we determine which beliefs and desires to attribute to people is by observing their behavior and then attributing the intentional states that best explain their behavior. A second simulation-based strategy for determining which beliefs and desires to attribute focuses on the agent's perceptual situation and on his or her "basic likings or cravings." [Goldman 1989, p. 170]

> From your perceptual situation, I infer that you have certain perceptual experiences or beliefs, the same ones I would have in your situation. I may also assume (pending information to the contrary) that you have the same basic likings that I have: for food, love, warmth, and so on. [Goldman 1989, p. 170]

As we read them, there is only one important point on which Gordon and Goldman actually *dis*agree. The accounts of simulation-based prediction, explanation and interpretation that we have sketched all seem to require that the person doing the simulating must already understand intentional notions like belief and desire. A person can't pretend he believes that the cookies are in the cookie jar unless he understands what it is to believe that the cookies are in the cookie jar; nor can a person imagine that she wants to make her friend laugh unless she understands what it is to want to make someone laugh. Moreover, as Goldman notes, when simulation is used to attribute intentional states to agents, it "assumes a prior understanding of what state it is that the interpreter attributes to [the agent]." [Goldman 1989, p. 182] Can the process of simulation somehow be used to explain the meaning or truth conditions of locutions like "S believes that p" and "S desires that q"? Goldman is skeptical, and tells us that "the simulation theory looks distinctly unpromising on this score." [Goldman 1989, p. 182] But Gordon is much more sanguine. Building on earlier suggestions by Quine, Davidson and Stich, he proposes the following account:

> My suggestion is that
>
> > (2) [*Smith believes that* Dewey won the election.]
>
> be read as saying the same thing as
>
> > (1) [Let's do a Smith simulation. Ready? *Dewey won the election*.]
>
> though less explicitly. [Gordon 1986, p. 167]

We are not at all sure we understand this proposal, and Gordon himself concedes that "the exposition and defense of this account of belief are much in need of further development." [Gordon 1986, p. 167] But no matter. We think we do understand the simulation-based accounts of prediction, explanation and interpretation that Gordon and Goldman both endorse. We're also pretty certain that none of these accounts is correct. In the sections that follow, we will try to say why.

4. Arguments in Support of Simulation-Based Accounts

In this Section we propose to assemble all the arguments we've been able to find in favor of simulation-based accounts and say why we don't think any of them is persuasive. Then, in the following Section, we will go on to offer some arguments of our own aimed at showing that there is lots of evidence that simulation-based accounts cannot easily accommodate, though more traditional theory-based accounts can. Before turning to the arguments, however, we would do well to get a bit clearer about the questions that the arguments are (and are not) intended to answer.

The central idea in the accounts offered by Gordon and Goldman is that in predicting, explaining or interpreting other people we simulate them by using part of *our own* cognitive systems "off-line". There might, of course, be other kinds of simulation in which we do not exploit our own decision making system in order to model the person we are simulating. But these other sorts of simulation are not our current concern. To avoid confusion, we will henceforth use the term off-line simulation for the sort of simulation that Gordon and Goldman propose. The question in dispute, then, is whether off-line simulation plays a central role in predicting, explaining or interpreting other people. Gordon and Goldman say yes; we say no.

It would appear that the only serious alternatives to the off-line simulation story are various versions of the "theory-theory" which maintain that prediction, explanation and interpretation exploit an internally represented theory or knowledge structure - a tacitly known "folk psychology." So if an advocate of off-line simulation can mount convincing arguments against the theory-theory, then he can reasonably claim to have made his case. The theory-theory is not the only game in town, but it is the only *other* game in town. It is not surprising, then, that in defending off-line simulation Gordon and Goldman spend a fair amount of time raising objections to the theory-theory.

There are, however, some important distinctions to be drawn among different types of theory-theories. Until fairly recently, most models that aimed at explaining cognitive capacities posited internally represented knowledge structures that invoked explicit rules or explicit sentence-like principles. But during the last decade there has been a growing dissatisfaction with sentence-based and rule-based knowledge structures, and a variety of alternatives have been explored. Perhaps the most widely discussed alternatives are connectionist models in which the knowledge used in making predictions is stored in the connection strengths between the nodes of a network. In many of these systems it is difficult or impossible to view the network as encoding a set of sentences or rules. [Ramsey, Stich & Garon 1990] Other theorists have proposed quite different ways in which non-sentential and non-rule-like strategies could be used to encode information. [See, for example, Johnson-Laird 1983]

Unfortunately, there is no terminological consensus in this domain. Some writers prefer to reserve the term "theory" for sentence-like or rule-based systems. For these writers, most connectionist models do not invoke what they would call an internally represented theory. Other writers are more liberal in their use of "theory," and are prepared to count just about any internally stored body of information about a given domain as an internally represented theory of that domain. For these writers, connectionist models and other non-sentential models do encode a tacit theory. We don't think there is any substantive issue at stake here. But the terminological disagreements can generate a certain amount of confusion. Thus, for example, someone who used "theory" in the more restrictive way might well conclude that if a connectionist (or some other non-sentence based) account of our ability to predict other people's behavior turns out to be the right one, then the theory-theory is mistaken. So far, so good. But it is important to see that the falsity of the theory-theory (narrowly construed) is no comfort at all to the off-line simulation theorist. The choice between off-line simulation theories and theory-theories is plausibly viewed as exhaustive only when "theory" is used in the *wide* rather than the restrictive way. For the remainder of this paper, we propose to adopt the wide interpretation of "theory". Using this terminology, the geography of the options confronting us are represented in Figure 2.[(7)] In the pages that follow, we will be defending option (A) in answer to Question (I). We take no stand at all on question (II). So much for getting clear on the questions. Now let's turn to the arguments.

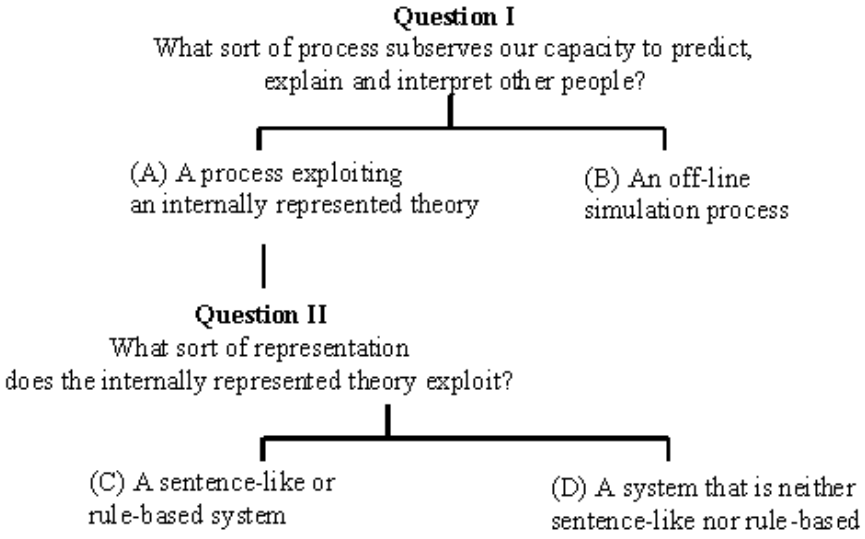

Figure 2

Argument 1: No one has been able to state the principles of the internally represented folk psychological theory posited by the theory-theory.

Both Goldman and Gordon go on at some length about the fact that it has proven very difficult to state the principles or laws of the folk psychological theory that, according to the theory-theorist, guide our interpretations and predictions.

> [A]ttempts by philosophers to articulate the putative laws or `platitudes' that comprise our folk theory have been notably weak. Actual illustrations of such laws are sparse in number; and when examples are adduced, they commonly suffer from one of two defects: vagueness and inaccuracy.... But why, one wonders, should it be so difficult to articulate laws if we appeal to them all the time in our interpretative practice? [Goldman 1989, p. 167; see also Gordon 1986, p. 166 & Gordon forthcoming, a, Sec. 3.7]

Reply: Goldman is certainly right about one thing. It is indeed very difficult to articulate the principles of folk psychology precisely and accurately. But it is hard to see why this fact should be of any comfort to advocates of the off-line simulation theory. For much the same could be said about the knowledge structures underlying all sorts of cognitive capacities. It has proven enormously difficult to state the principles underlying a speaker's capacity to judge the grammaticality of sentences in his language. Indeed, after three decades of sustained effort, we don't have a good grammar for even a single natural language. Nor do we have a good account of the principles underlying people's everyday judgements about the behavior of middle sized physical objects, or about their ability to solve mathematical problems, or about their ability to play chess, etc. But, of course, in all of these domains, the theory-theory really is the only game in town. The off-line simulation story makes no sense as an account of our ability to judge grammaticality, or of our ability to predict the behavior of projectiles.

The difficulties encountered by those who have sought to describe the rules or principles underlying our grammatical (or mathematical or physical) abilities have convinced a growing number of theorists that our knowledge in these domains is not stored in the form of rules or principles. That conviction has been an important motive for the development of connectionist and other sorts of non-sentential and non-rule based models. But none of this should encourage an advocate of the off-line simulation theory. The dispute between connectionist models and rule-based models is the dispute between

(C) and (D) in Figure 2.

And that is a dispute *among theory-theorists*. Of course on a narrow interpretation of "theory," on which only rule-based and sentence-based models count as theories, the success of connectionism would indeed show that the "theory-theory" is mistaken. But, as we have taken pains to note, a refutation of the theory-theory will support the off-line simulation account only when "theory" is interpreted broadly.

Argument 2: Mental simulation models have been used with some success by a number of cognitive scientists.

Here's how Goldman makes the point:

> [S]everal cognitive scientists have recently endorsed the idea of mental simulation as one cognitive heuristic, although these researchers stress its use for knowledge in general, not specifically knowledge of others' mental states. Kahneman and Tversky 1982 propose that people often try to answer questions about the world by an operation that resembles the running of a simulation model. The starting conditions for a `run', they say, can either be left at realistic default values or modified to assume some special contingency. Similarly, Rumelhart [et. al.] describe the importance of `mental models' of the world, in particular, models that simulate how the world would respond to one's hypothetical actions. [Goldman 1989, p. 174]

Reply: Here, again, it is our suspicion that ambiguity between the two interpretations of "theory" is lurking in the background and leading to mischief. The "simulation" models that Goldman cites are the sort that would be classified under (D) in Figure 2. If they are used in the best explanation of a given cognitive capacity, then that capacity is subserved by a tacit theory, and *not* by an off-line simulation. Of course when "theory" is read narrowly, this sort of simulation will not count as a tacit theory. But, as already noted, on the narrow reading of "theory" the falsity of internalized theory accounts lends no support at all to the off-line simulation theory.

Argument 3: "To apply the alleged common-sense theory would demand anomalous precocity."

What we've just quoted is a section heading in one of Gordon's unpublished paper.[(8)] He goes on to note that recent studies have shown children as young as two and a half "already see behavior as dependent on belief and desire." It is, he suggests, more than a bit implausible that children this young could acquire and use "a theory as complex and sophisticated" as the one that the theory-theory attributes to them. Goldman elaborates the argument as follows:

> [C]hildren seem to display interpretive skills by the age of four, five or six. If interpretation is indeed guided by laws of folk psychology, the latter must be known (or believed) by this age. Are such children sophisticated enough to represent such principles? And how, exactly, would they acquire them? One possible mode of acquisition is cultural transmission (e.g. being taught them explicitly by their elders). This is clearly out of the question, though, since only philosophers have even tried to articulate the laws, and most children have no exposure to philosophers. Another possible mode of acquisition is private construction. Each child constructs the generalizations for herself, perhaps taking clues from verbal explanations of behavior which she hears. But if this construction is supposed to occur along the lines of familiar modes of scientific theory construction, some anomalous things must take place. For one thing, all children miraculously construct the same nomological principles. This is what the (folk-) theory theory ostensibly implies, since it imputes a single folk psychology to everyone. In normal cases of hypothesis construction, however, different scientists come up with different theories. [Goldman 1989, pp. 167-8]

Reply: There is no doubt that if the theory-theory is right, then the child's feat is indeed an impressive one. Moreover, it is implausible to suppose that the swift acquisition of folk psychology is subserved by the same learning mechanism that the child uses to learn history or chemistry or astronomy. But, once again, we find it hard to see how this can be taken as an argument against the theory-theory and in favor of the off-line simulation theory. For there are other cases in which the child's accomplishment is comparably impressive and comparably swift. If contemporary generative grammar is even *close* to being right, the knowledge structures that underlie a child's linguistic ability are enormously complex. Yet children seem to acquire the relevant knowledge structures even more quickly than they acquire their knowledge of folk psychology. Moreover, children in the same linguistic community all acquire much the same grammar, despite being exposed to significantly different samples of what will become their native language. Less is known about the knowledge structures underlying children's abilities to anticipate the behavior of middle sized physical objects. But there is every reason to suppose that this "folk physics" is at least as complex as folk psychology, and that it is acquired with comparable speed. Given the importance of all three knowledge domains, it is plausible to suppose that natural selection has provided the child with lots of help - either in the form of innate knowledge structures or in the form of special purpose learning mechanisms. But whatever the right story about acquisition turns out to be, it is perfectly clear that in the case of grammar,

and in the case of folk physics, what is acquired must be some sort of internally represented theory. Off-line simulation could not possibly account for our skills in those domains. Since the speed of language acquisition and the complexity of the knowledge acquired do not (indeed, could not) support an off-line simulation account of linguistic ability, we fail to see why Gordon and Goldman think that considerations of speed and complexity lend any support at all to the off-line simulation account of our skills in predicting, explaining and interpreting behavior.

Argument 4: The off-line simulation theory is much simpler than the theory-theory.

Other things being equal, we should surely prefer a simpler theory to a more complex one. And on Gordon's view,

> the simulation alternative makes [the theory-theory] strikingly unparsimonious. Insofar as the store of causal generalizations posited by [the theory-theory] mirrors the set of rules *our own* thinking typically conforms to, the Simulation Theory renders it altogether otiose. For whatever rules our own thinking typically conforms to, our thinking continues to conform to them within the context of simulation.... In the light of this far simpler alternative, the hypothesis that people must be endowed with a special stock of laws corresponding to rules of logic and reasoning is unmotivated and unparsimonious. [Gordon forthcoming, a, Sec. 3, p. 7]

Reply: When comparing the simplicity of a pair of theories, it is important to look at the whole theory in both cases, not just at isolated parts. It is our contention that if one takes this broader perspective, the greater parsimony of the simulation theory simply disappears. To see the point, note that for both the theory-theory and the simulation theory the mechanism subserving our predictions of other people's behavior must have two components. One of these may be thought of as a data base that somehow stores or embodies information about how people behave. The other component is a mechanism which applies that information to the case at hand - it extracts the relevant facts from the data base. Now if we look only at the data base, it does indeed seem that the theory-theory is "strikingly unparsimonious" since it must posit an elaborate system of internally represented generalizations or rules - or perhaps some other format for encoding the regularities of folk psychology. The simulation theory, by contrast, uses the mind's decision making system as its "data base," and that decision making system would have to be there on any theory, because it explains how we make real, "on-line" decisions. So the off-line simulation theory gets its data base for free.

But now let's consider the other component of the competing theories. Merely *having* a decision making system will not enable us to make predictions about other people's behavior. We also need the capacity to take that system "off-line," feed it "pretend" inputs and interpret its outputs as predictions about how someone else would behave. When we add the required cognitive apparatus, the picture of the mind that emerges is sketched in Figure 3. Getting this "control mechanism" to work smoothly is sure to be a *very* non-trivial task. How do things look in the case of the theory-theory? Well, no matter how we go about making predictions about other people, it is clear that in making predictions about physical systems we can't use the off-line simulation strategy; we have to use some sort of internalized theory (though, of course, it need not be a sentence-like or rule-based theory). Thus we know that the mind is going to have to have some mechanism for extracting information from internalized theories and applying it to particular cases. (In Figure 1 we have assumed that this mechanism is housed along with the other "Inference Mechanisms" that are used to extract information from pre-existing beliefs.) If such a mechanism will work for an internally represented folk physics, it is plausible to suppose that, with minor modifications, it will also work for an internally represented folk psychology. So while the simulation theorist gets the data base for free, it looks like the theory-theorist gets the "control mechanism" for free. All of this is a bit fast and loose, of course. But we don't think either side of this argument can get much more precise until we are presented with up-and-running models to compare. Until then, neither side can gain much advantage by appealing to simplicity.

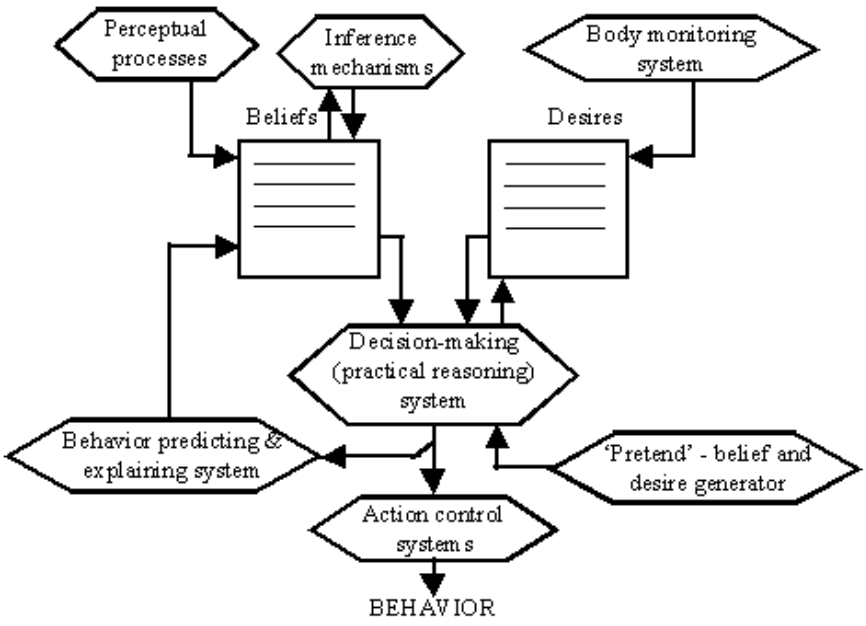

Figure 3

Argument 5: When we introspect about our predictions of other people's behavior, it sometimes seems that we proceed by imagining how we would behave in their situation.

Here is how Goldman makes the point:

> The simulation idea has obvious initial attractions. Introspectively, it seems as if we often try to predict others' behavior - or predict their (mental) choices - by imagining ourselves in their shoes and determining what we would choose to do. [Goldman 1989, p. 169]

And here is Gordon:

> [C]hess players report that, playing against a human opponent or even against a computer, they visualize the board from the other side, taking the opposing pieces for their own and vice versa. Further, they pretend that their reasons for action have shifted accordingly.... Thus transported in imagination, they `make up their mind what to do.' [Gordon 1986, pp. 161-2]

Both authors are aware that appeal to introspection can be a two edged sword, since it also often happens that we predict other people's behavior *without* introspecting any imaginary behavior.

> [T]here is a straightforward challenge to the psychological plausibility of the simulation approach. It is far from obvious, introspectively, that we regularly place ourselves in another person's shoes, and vividly envision what we would do in his circumstances. [Goldman 1989, p. 176]

> Imagery is not always needed in such simulations. For example, I need no imagery to simulate having a million dollars in the bank. [Gordon 1986, p. 161]

To deal with this "challenge," Goldman proposes a pair of replies. First, simulation need not always be introspectively

vivid. It can often be "semi-automatic, with relatively little salient phenomenology." [Goldman 1986, p. 176] Second, not all interpretations rely on simulation. In many cases interpreters rely solely on "inductively acquired information" though the information is "historically derived from earlier simulations." [Goldman 1986, p. 176]

Reply: We don't propose to make any fuss at all about the frequent absence of "salient phenomenology." For it is our contention that when the issue at hand is the nature of the cognitive mechanism subserving our capacity to interpret and predict other people's behavior, the entire issue of introspective imagination is a red herring. Indeed, it is *two* red herrings. To see the first of them, consider one of the standard examples used to illustrate the role of imagery in thought. Suppose we ask you: "How many windows are there in your house?" How do you go about answering? Almost everyone reports that they *imagine* themselves walking from room to room, counting the windows as they go. What follows from this about the cognitive mechanism that they are exploiting? Well, one thing that surely *does not* follow is that off-line simulation is involved. The *only* way that people could possibly answer the question accurately is to tap into some internally represented store of knowledge about their house; it simply makes no sense to suppose that off-line simulation is being used here. So even if a cognitive process is *always* accompanied by vivid imagery, that is no reason at all to suppose that the process exploits off-line simulation. From this we draw the obvious conclusion. The fact that prediction and interpretation *sometimes* involve imagining oneself in the other person's shoes is less than no reason at all to suppose that off-line simulation is involved.

It might be suggested that, though imagery provides no support for the off-line simulation hypothesis, it does challenge the theory-theory when "theory" is interpreted narrowly. For it shows that some of the information we are exploiting in interpretation and prediction is not stored in the form of sentences or rules. But even this is far from obvious. There is a lively debate in the imagery literature in which "descriptionalists," like Pylyshyn and Dennett, maintain that the mechanisms underlying mental imagery exploit language-like representations, while "pictorialists," like Kosslyn and Fodor, argue that images are subserved by a separate, non-linguistic sort of representation. [Pylyshyn 1981, Dennett 1969 & 1978, Fodor 1975, Kosslyn 1981] We don't propose to take sides in this dispute. For present purposes it is sufficient to note that, unless it is supplemented by a persuasive argument in favor of pictorialism and against descriptionism, the introspective evidence does not even challenge the theory-theory *construed narrowly*.

Argument 6: The off-line simulation account is supported by recent experimental studies focusing on children's acquisition of the ability to interpret and predict other people.

On our view, this is far and away the most interesting argument that has been offered in favor of the off-line simulation theory. To see exactly what the experimental studies do, and do not, support, we'll have to look at both the evidence and the argument with considerable care. Gordon does a good job of describing one important set of experiments.

> Very young children give verbal expression to predictions and explanations of the behavior of others. Yet up to about the age of four they evidently lack the concept of belief, or at least the capacity to make allowances for false or differing beliefs. Evidence of this can be teased out by presenting children with stories and dramatizations that involve dramatic irony: where we the audience know something important the protagonist doesn't know....
>
> In one such story (illustrated with puppets) the puppet-child Maxi puts his chocolate in the box and goes out to play. While he is out, his mother transfers the chocolate to the cupboard. Where will Maxi look for the chocolate when he comes back? In the box, says the five year old, pointing to the miniature box on the puppet stage: a good prediction of a sort we ordinarily take for granted.... But the child of three to four years has a different response: verbally or by pointing, the child indicates the cupboard. (That is, after all, where the chocolate is to be found, isn't it?) Suppose Maxi wants to mislead his gluttonous big brother to the *wrong* place, where will he lead him? The five year old indicates the cupboard, where (unbeknownst to Maxi) the chocolate actually is.... The *younger* child indicates, incorrectly, the box. [Gordon 1986, p. 168]

These results, Gordon maintains, are hard to square with the theory-theory. For if the theory-theory is correct, then

> before internalizing [the laws and generalizations of folk psychology] the child would simply be unable to predict or explain human action. And *after* internalizing the system, the child could deal indifferently with actions caused by *true* beliefs and actions caused by *false* beliefs. It is hard to see how the semantical question could be relevant. [Gordon 1986, p. 169]

But, according to Gordon, these data are just what we should expect, if the off-line simulation theory is correct.

> The Simulation Theory [predicts that] prior to developing the capacity to simulate others for purposes of

> prediction and explanation, a child will make *egocentric errors* in predicting and explaining the actions of others. She will predict and explain as if whatever she herself counts as "fact" were also fact to the other; which is to say, she fails to make allowances in her predictions and explanations for false beliefs or for what the other isn't in a position to know. [Gordon unpublished, a, Sec. 3.6, p. 11]

Reply: According to Gordon, the theory-theory can't easily explain the results of the "Maxi" experiment, though the off-line simulation theory predicts those results. We're not convinced on either score. Let's look first at just what the off-line simulation story would lead us to expect.

Presumably by the time any of these experiments can be conducted, the child has developed a more or less intact decision making system like the one depicted in Figure 1. That system makes "on-line" decisions and thus determines the child's actions on the basis of her actual beliefs and desires. But by itself it provides the child with no way of predicting Maxi's behavior or anyone else's. If the off-line simulation theory is right, then in order to make predictions about other people's behavior two things must happen. First, the child must acquire the ability to take the output of the decision making system off-line - treating its decisions as predictions or expectations, rather than simply feeding them into the action controlling system. Second, the child must acquire the ability to provide the system with input other than her own actual beliefs and desires. She must be able to supply the system with "pretend" input so that she can predict the behavior of someone whose beliefs and desires are different from her own. (These are the two capacities that are represented in Figure 3 and absent in Figure 1.) There is, of course, no *a priori* reason to suppose that these two steps happen at different times, nor that the one we've listed first will occur first. But if they do occur in that order, then we might expect there to be a period when the child could predict her own behavior (or the behavior of someone whose beliefs and desires are the same as hers) though she could not predict the behavior of people whose beliefs or desires are different from hers. It is less clear what to expect if the steps occur in the opposite order. Perhaps the result would be some sort of pretending or play-acting - behaving in a way that someone with different beliefs or desires would behave. Though until the child develops the capacity to take output of the decision making system off-line, she will not be able to predict other people's behavior or her own. So it looks like the off-line simulation story makes room for three possible developmental scenarios.

> (1) The child acquires both abilities at the same time. In this case we would expect to see two developmental stages. In the first the child can make no predictions. In the second she can make a full range of predictions about people whose beliefs and desires are different from her own.
>
> (2) The child first acquires the ability to take the output off-line, and then acquires the ability to provide the system with pretend input. In this case we would expect three developmental stages. In the first, the child can make no predictions. In the second, she can only make predictions about her own behavior or about the behavior of people whose beliefs and desires are identical to hers. In the third, she can make the full range of predictions.
>
> (3) The child first acquires the ability to provide the system with pretend inputs, and then acquires the ability to take the output off-line. In this case, too, we would expect three developmental stages. The first and last stages are the same as those in (2), but in the middle stage the child can play-act but not make predictions.

Now let's return to the Maxi experiment. Which of these developmental scenarios do the children in these experiments exhibit? At first blush, it might be thought that the pattern Gordon reports is much the same as the one set out in scenario (2). But that would be a mistake. The younger children - those who are giving the wrong answers - are not predicting that Maxi would do what someone with their own beliefs and desires would do. For they have no desire to get the chocolate, nor to deceive the gluttonous brother. Those are *Maxi's* desires, not *theirs*. If anything, it would appear that these children are half way between the second and third stages of scenario (2): they can feed "pretend" desires into the decision making system, but not "pretend" beliefs. Of course none of this shows that the off-line simulation theory is false. It is perfectly compatible with the theory to suppose that development proceeds as in (2), *and* that the transition from the second to the third stage proceeds in two sub-stages - desires first, and then beliefs. (This pattern is sketched in Figure 4.) But it is, to say the least, something of an exaggeration to say that the off-line simulation theory "predicts" the experimental results. The most that can be said is that the theory is compatible with the observed developmental pattern, and with lots of other patterns as well.[(9)]

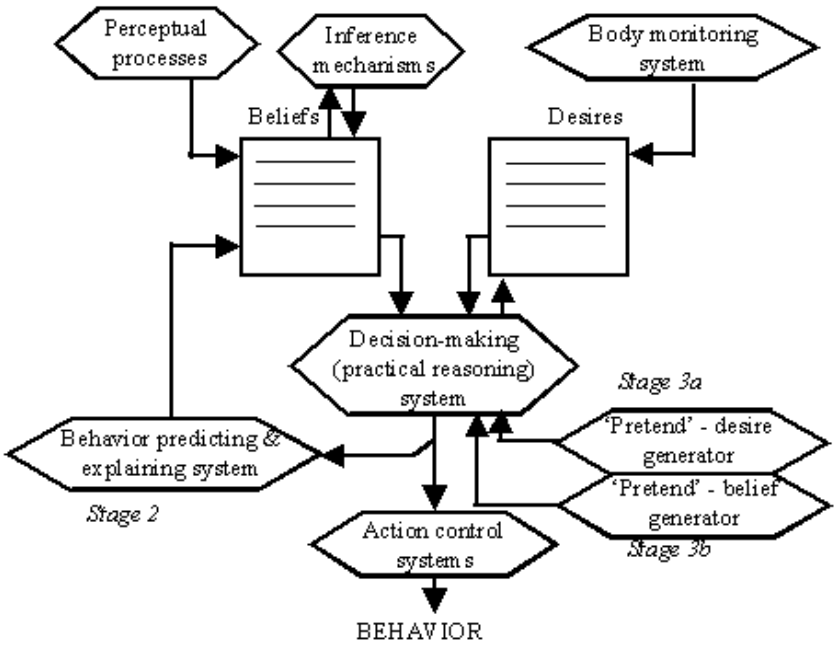

Figure 4

For the results that Gordon describes to be at all relevant to the dispute between the off-line simulation theory and the theory-theory it would have to be the case that the latter theory is *in*compatible with the reported developmental pattern. But that is patently not the case. To see why, we should first note that the theory-theory is not committed to the claim that folk psychology is acquired all in one fell swoop. Indeed, one would expect just the opposite. If children really are acquiring a tacit theory of the mind, they probably acquire it a bit at a time. Thus it might be the case that, at a given stage in development, children have mastered the part of the theory that specifies how beliefs and desires lead to behavior, though they have not mastered the entire story about how beliefs are caused. At this stage, they might simply assume that beliefs are caused by the way the world is; they might adopt the strategy of attributing to everyone the very same beliefs that they have. A child who has acquired this much of folk psychology would (incorrectly) attribute to Maxi the belief that the chocolate is in the cupboard. She would then go on to make just the predictions that Gordon reports. Of course, the theory-theory is also compatible with lots of other hypotheses about which bits of folk psychology are acquired first. Thus, like the off-line simulation theory, it is compatible with (but does not entail) lots of possible developmental patterns. So it looks like the developmental studies that Gordon and Goldman cite can't be used to support one theory over the other.

Argument 7: Autistic children are highly deficient in their ability to engage in pretend play. These children are also frequently unable to impute beliefs to others or to predict other people's behavior correctly.

Here's how Gordon sets out the argument:

> Practical simulation involves the capacity for a certain kind of systematic pretending. It is well known that *autistic* children suffer a striking deficit in the capacity for pretend-play. In addition, they are often said to `treat people and objects alike'; they fail to treat others as subjects, as having `points of view' distinct from their own. This failure is confirmed by their performance in prediction tests like the [Wimmer-Perner "Maxi" experiment] I have just described. A version of the Wimmer-Perner test was administered to autistic children of ages *six to sixteen* by a team of psychologists.... *Almost all* these children gave the wrong answer, the 3-year-old's answer. This indicates a highly specific deficit, not one in general intelligence. Although many

> autistic children are also mentally retarded, those tested were mostly in the average or borderline IQ range. Yet children with Down's syndrome, with IQ levels substantially below that range, suffered no deficit: almost all gave the right answer. My account of belief would predict that only those children who can engage in pretend play can master the concept of belief. [Gordon 1986, p. 196]

Goldman is rather more tentative. He claims only that the inability of autistic children "to impute beliefs to others and therefore predict their behavior correctly .... might ... be related to their lack of pretend play." [Goldman 1989, p. 175]

Reply: The fact that autistic children are both incapable of pretend play and unable to predict the behavior of other people in Wimmer-Perner tests is very intriguing. Moreover, Gordon is certainly right in suggesting that the off-line simulation theory provides a possible explanation for these facts. If the off-line simulation theory is right, predicting the behavior of people whose beliefs differ from our own requires an ability to provide our own decision making system with pretend input. And it is plausible to assume that this ability would also play a central role in pretend play. So if we hypothesize that autistic children lack the ability to provide the decision making system with pretend input, we could explain both their performance on the Wimmer-Perner test and their failure to engage in pretend play. But, of course, this will not count as an argument for the off-line simulation theory and against the theory-theory if the latter account can offer an equally plausible explanation of the facts. And it will require no creativity on our part to produce such an alternative explanation since one of the investigators who discovered the fact that autistic children do poorly in Wimmer-Perner tests has offered one himself.

Leslie (1988) takes as an assumption "the hypothesis that human cognition involves *symbolic computations* in the sense discussed ... by Newell (1980) and particularly by Fodor." [Leslie 1988, p. 21] He also assumes that an internalized theory of mind underlies the normal adult's ability to predict other people's behavior. An important theme in Leslie's work is that developmental studies with both normal and autistic children can help to illuminate the expressive resources of the "language of thought" in which our theory of mind is encoded. According to Leslie, the notion of a "meta-representation" is central in understanding how our theory of mind develops. Roughly speaking, a meta-representation is a mental representation about some other representational state or process. We exploit meta-representations when we think that

> Maxi believes that the chocolate is in the box

or that

> Maxi's brother wants the chocolate

or that

> Mommy is pretending that the banana is a telephone.

On Leslie's view, "autistic children do not develop a theory of mind normally." [Leslie 1988, p. 39] And while "it is far too soon to say with any confidence what *is* wrong" with these children, he speculates that at the root of the problem may be an inability to use meta-representations. If this were true, it would explain both their difficulty with pretend play and their failure on the Wimmer-Perner test.

Though we find Leslie's speculation interesting and important, it is no part of our current project to defend it. To make our case we need only insist that, on currently available evidence, Leslie's hypothesis is no less plausible than Gordon's. Since Leslie's speculation presupposes that normal children acquire and exploit a theory of mind that is encoded in a language of thought, the evidence from studies of autistic children gives us no reason to prefer the off-line simulation account over the theory-theory.(10)

Our theme, in this reply and in the previous one, has been that the empirical evidence cited by Gordon and Goldman, while compatible with the off-line simulation theory, is also compatible with the theory-theory, and thus does not support one theory over the other. But there are other studies in the recent literature that *can* be used to support one theory over the other. These studies report results that are comfortably compatible with the theory-theory though not with the off-line simulation account. Before we sketch those results, however, it is time to start a new section. In this section we've tried to show that none of the arguments in favor of the off-line simulation theory are persuasive. In the next one we'll set out a positive case for the theory-theory.

5. In Defense of the Theory-Theory

Argument 1: There are developmental data that are easily accommodated by the theory-theory, but very hard to explain if

the off-line simulation account is correct.

Let's start with a description of the experimental setup, and a quick overview of the data.

> The setup of the task in these experiments was rather simple. Two children were placed facing each other on opposite sides of a table. In each trial one child served as subject and had access to the other child's knowledge and his or her own knowledge of the content of a closed box. The box was placed in the middle of the table between the two children. The outside of it was neutral and not suggestive of its content. In each box was a familiar object like a pencil, a comb, a piece of chocolate, and so on. The specific questions were: "Does (name of other child) know what is in the box or does she not know that?" and "Do you know what is in the box or do you not know that?"....
>
> Before the knowledge-questions were asked, either the other or the subject had access to the content of the box. One kind of access was visual perception. In this case either the other child or the subject had a chance to look into the box. The other kind of access was verbal information. Here the experimenter looked into the box and then informed one of the children by whispering the name of the content object into the child's ear. Because the two children were facing each other the subject was fully aware of the informational conditions the other child was exposed to, that is, of whether the other child did or did not look into the box and of whether the other was or was not informed. [Wimmer, et. al. 1988, p. 175]

The results of this experiment were quite striking. The older children (5-year-olds) gave uniformly correct answers. But younger children (3- and 4-year-olds) did not.

> The most frequent error was denial of the other child's knowledge when the other child had looked into the box or was informed by the experimenter.
>
> Most 3-year-olds and some 4-year-olds said that the other did not know what was in the box. This kind of error was nearly absent in children's assessment of their own knowledge. When subjects themselves had looked into the box or were informed, then they claimed to know and they could, of course, tell what was in the box. [Wimmer, et. al. 1988, pp. 175-6]

In another experiment, designed to be sure that the younger children were aware the other child had looked in the box, the subjects were asked both whether the other child had looked in the box and whether the other child knew what was in the box. "The children consistently responded affirmatively to the look-question but again quite frequently responded negatively to the knowledge question." [Wimmer, et. al. 1988, p. 176]

What is going on here? The explanation offered by the experimenters is that younger children are using quite different mental processes in assessing what they know and in assessing what the other child knows. To answer the question. "Do you know what is in the box?" the children use what the experimenters call the "answer check procedure." They simply check to see whether they have an answer to the embedded question in their knowledge base, and if they do they respond affirmatively. To answer the question about the other child's knowledge, the older children used what the experimenters call a "direct access check procedure". In effect, they ask themselves whether the other child looked in the box or was told about its contents. If so, they respond affirmatively. If not, they respond negatively. However, the 3-year-olds did not use this procedure. They simply checked whether the other child had uttered a correct statement about the box's contents. If she had not, the subject said the other did not know. A very natural way to describe the situation is that while the younger children know that people who say *that p* typically believe or know *that p*, these children have not yet learned that people will come to know *that p* by seeing or being told *that p*. The younger children have acquired a fragment of folk psychology, while the older children have acquired a more substantial piece of the theory.[(11)] The older children have not, however, entirely mastered the theory, as indicated by another series of experiments.

These experiments focused on the role of *inference* in the acquisition of knowledge or belief. What they show is that "four- and 5-year olds relied on inference in their own acquisition of knowledge but denied that the other person might know via inference." [Wimmer, et. al. 1988, p. 179]

> Inferential access was realized in these experiments in a very simple and concrete way. In a first step the child and the other person together inspected the content of a container and agreed that only sweets of a certain kind, for example, black chocolate nuts, were in the container. In a second step the other person or the subject was prevented from seeing how one choconut was transferred from the container into an opaque bag. However, this person was explicitly informed by the experimenter about this transfer, for example, "I've just taken one of the things out of this box and put it in the bag."

> The condition where knowledge could be acquired via simple inference was contrasted with a condition where knowledge depended on actually seeing the critical object's transfer. In this latter condition two kinds of sweets were in the original container, and thus one could only know what the content of the critical bag was by having seen the transfer from container to bag. [Wimmer, et. al. 1988, p. 179]

Once again, the results were quite striking. In most cases the older children (6-year-olds, in this case) generally gave the right answers both about their own knowledge and about the other child's knowledge. But although the 4-year-olds used inference in forming their own beliefs, a substantial majority of them exhibited a pattern that the experimenters called "inference neglect."

> The response pattern "inference neglect" means that the other person was assessed according to perceptual access: When the other person saw the object's transfer to the bag, 4-year-olds attributed knowledge; when the other did not see this transfer, ignorance was attributed even when the other person in fact knew via inference. [Wimmer, et. al. 1988, p. 179]

One plausible way of accounting for these results is to hypothesize that the older children had mastered yet another part of the adult folk psychology. They had learned that knowledge and beliefs can be caused by inference as well as by direct perceptual access. And, indeed, this is just the interpretation that the experimenters suggest.

> In contrast to the 3-year-olds discussed in the previous [experiment], the 4- and 5-year-olds in the present experiments understood quite well that one has to consider the other person's informational conditions when one is questioned about the other person's knowledge. Their only problem was their limited understanding of informational conditions. They understood only direct visual access as a source of knowledge and this led them to mistaken but systematic ignorance attributions in the case of inferential access. [Wimmer, et. al. 1988, p. 181]

Let's now ask what conclusions can be drawn from these experiments that will be relevant to the choice between the off-line simulation theory and the theory-theory. A first obvious fact is that the data are all comfortably compatible with the theory-theory. Indeed, the explanation of the data offered by the experimenters is one that presupposes the correctness of the theory-theory. What appears to be happening is that as children get older, they master more and more of the principles of folk psychology. By itself, of course, the theory-theory would not enable us to predict the data, since the theory-theory does not tell us anything about the order in which the principles of folk psychology are acquired. But the pattern of results described certainly poses no problem for the theory-theory.

The same cannot be said for the off-line simulation theory. It is clear that even the younger children in these studies form beliefs as the result of perception, verbally provided information, and inference. So there is nothing about their decision making system, when it is being used on-line, that will help to explain the results. To make predictions about other people, the off-line simulation theory maintains, children must acquire the capacity to take the decision making system off-line and provide it with some pretend inputs. But there is no obvious way in which this process could produce the pattern of results that has been reported. The difficulty is particularly clear in the case of inference. If the subject has seen that the box contains only chocolate nuts, and if she is told that one of the items in the box has been put in the bag, she comes to believe that there is a chocolate nut in the bag. But if she knows the other child has also seen what is in the box, and that the other child has been told that one of the items in the box has been put in the bag, she insists that the other child does not know what is in the bag. The problem can't be that the subject doesn't realize that the other child knows what is in the box. Children of this age do a good job of attributing belief on the basis of perception. Nor can it be that the subject doesn't believe that the other child believes the transfer has been made. For children of this age are also adept at attributing beliefs on the basis of verbally communicated information. So it looks like the subject has all the information needed for a successful simulation. But the answer she comes up with is ***not*** the one that she herself would come up with, were she in the subject's place. There are, of course, endlessly many ways in which a resolute defender of the off-line simulation theory might try to accommodate these data. But all the ones we've been able to think of are obviously implausible and ad hoc.

Argument 2: Our predictions and explanations of behavior are "cognitively penetrable."

One virtue of using a simulation to predict the behavior of a system is that you need have no serious idea about the principles governing the behavior of the target system. You just run the simulation and watch what happens. Sometimes, of course, a simulation will do something that was utterly unexpected. But no matter. If the simulation really was similar to the target system, then the prediction it provides will be a good one. In predictions based on simulations, what you don't know won't hurt you. All of this applies to the off-line simulation theory, of course. If there is some quirk in the human decision making system, something quite unknown to most people that leads the system to behave in an unexpected way under certain circumstances, the accuracy of predictions based on simulations should not be adversely affected. If you

provide the system with the right pretend input, it should simulate (and thus predict) the unexpected output. Adapting a term from Pylyshyn, we might describe this by saying that simulation-based predictions are not "cognitively penetrable." (12)

Just the opposite is true for predictions that rely on a theory. If we are making predictions on the basis of a set of laws or principles, and if there are some unexpected aspects of the system's behavior that are not captured by our principles, then our predictions about those aspects of the system's behavior should be less accurate. Theory based predictions are sensitive to what we know and don't know about the laws that govern the system; they *are* cognitively penetrable. This contrast provides a useful way of testing the two theories. If we can find cases in which ignorance about the workings of one's own psychology leads people to make mistakes in predicting what they, or other similarly situated people, would do, it will provide yet another reason to think that the off-line simulation theory is untenable. And, as it happens, cases illustrating cognitive penetrability in the prediction of behavior are not all that hard to find. The literature in cognitive social psychology is full of them. We'll illustrate the point with three examples, but it would be easy to add three dozen more.

First Example: Suppose you are walking through the local shopping mall, and encounter what looks to be yet another consumer product opinion survey. In this one a polite, well dressed man invites you to examine an array of familiar products - nightgowns, perhaps, or pantyhose - and to rate their quality. A small reward is offered for your participation - you can keep the garment you select. On examining the products, you find no really significant differences among them. (You couldn't, because, unbeknownst to you they are identical.) What would you do? Confronted with this question, most of us think we would report that the garments looked to be very similar, and then choose one randomly. However, when the experiment was actually tried, this turned out to be mistaken. "There was a pronounced position effect on evaluations, such that the right-most garments were heavily preferred to the left-most garments." But it was clear that few of the subjects had any awareness at all of the effect of position on their decision. Indeed, "when questioned about the effect of the garments' position on their choices, virtually all subjects denied such an influence (usually with a tone of annoyance or of concern for the experimenter's sanity)." [Nisbett & Ross 1980, p. 207]

This sort of case poses real problems for the off-line simulation theory. Most people have no trouble imagining themselves in the situation described. They can supply their decision making system with vivid "pretend" input. But few people who have not heard of the experiment predict that they would behave in the way that the subjects behaved. The natural interpretation of the experiment is that people's predictions about their own behavior (and the subjects' explanations of their own choice) are guided by an incomplete or inaccurate theory, one which includes no information about these so-called "position effects."

Second Example: Here's another case to run through your own simulator. Suppose someone in the office is selling $1.00 tickets for the office lottery. In some cases, when a person agrees to buy a ticket, he or she is simply handed one. In other cases, after agreeing to buy a ticket, the buyer is allowed to choose a ticket from several that the seller has available. On the morning of the lottery, the seller approaches each purchaser and attempts to buy back the ticket. Now imagine yourself in both roles - first as a person who had been handed the ticket, second as a person who had been given a choice. What price would you ask in each case? Would there be any difference between the two cases? On several occasions one of us (Stich) has asked large undergraduate classes to predict what they would do. Almost no one predicts that they would behave the way that people actually do behave. Almost everyone is surprised to hear the actual results.

Ah, yes, the results; we haven't yet told *you* what they are. When the experiment was actually done, "no-choice subjects sold their tickets back for an average of $1.96. Choice subjects, who had personally selected their tickets, held out for an average of $8.67!" [Nisbett & Ross 1980, p. 136] If, like Stich's students, you find this surprising and unexpected, it counts as yet another difficulty for the off-line simulation theory.

Third Example: In the psychology laboratory, and in everyday life, it sometimes happens that people are presented with fairly persuasive evidence that they have some hitherto unexpected trait. In the light of that evidence people form the belief that they have the trait. What will happen to that belief if, shortly after this, people are presented with a convincing case discrediting the first body of evidence? Suppose, for example, they are convinced that the test results were actually someone else's, or that no real test was conducted at all. Most people expect that the undermined belief will simply be discarded. If until recently I never had reason to think I had a certain trait, and if the evidence I just acquired has been soundly discredited, then surely it would be silly of me to go away thinking that I *do* have the trait. That seems to be what most people think. And the view was shared by a generation of social psychologists who duped subjects into believing all sorts of things about themselves, observed their reactions, and then "debriefed" the subjects by explaining the ruse. The assumption was that no enduring harm could be done because once the ruse was explained the induced belief would be discarded. But in a widely discussed series of experiments, Ross and his co-workers have demonstrated that this is simply not the case. Once a subject has been convinced that she is very good at telling real from fake suicide notes, for example, showing her that the evidence was completely phony does not succeed in eliminating the belief. Moreover, third person

observers of the experiment exhibit even stronger "belief perseverance." If an observer subject watches a participant subject being duped and then debriefed, the observer, too, will continue to believe that the participant is particularly good at detecting real suicide notes. [Nisbett & Ross 1980, p. 175-179]

Neither of these results should have been at all surprising to anyone if we predict each other's beliefs and behavior in the way that the off-line simulation theory suggests. But clearly the results *were* both surprising and disturbing. We can't simply ask ourselves what we would do in these circumstances and expect to come up with the right answer. For the theory-theorist, this fact poses no particular problem. When our folk psychology is wrong, it is to be expected that our predictions will be wrong too. It is simply another illustration of cognitive penetrability in predicting and explaining behavior. The theory-theory, unlike the off-line simulation theory, predicts that people's predictions and explanations of behavior will be cognitively penetrable through and through. If it is agreed that these experiments confirm cognitive penetrability, the off-line simulation theory is in serious trouble.

6. Conclusion

Our paper has been long but our conclusion will be brief. The off-line simulation theory poses an intriguing challenge to the dominant paradigm in contemporary cognitive science. Moreover, if it were correct the off-line simulation account of psychological prediction and explanation would largely undermine *both* sides in the eliminativism dispute. But it has been our contention that the prospects for the off-line simulation theory are not very bright. None of the arguments that have been offered in defense of the theory are at all persuasive. And there is lots of experimental evidence that would be very hard to explain if the off-line simulation account were correct. We don't claim to have provided a knock-down refutation of the off-line simulation theory. Knock-down arguments are hard to come by in cognitive science. But we do claim to have assembled a pretty serious case against the simulation theory. Pending a detailed response, we don't think the off-line simulation theory is one that cognitive scientists or philosophers should take seriously.[(13)]

REFERENCES

Astington, J., Harris, P., and Olson, D. (eds.) 1988. *Developing Theories of Mind*. Cambridge: Cambridge University Press.

Baron-Cohen, S., Leslie, A. and Frith, U. 1985. Does the Autistic Child Have a "Theory of Mind"? *Cognition*, 21, 37-46.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

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

Churchland, P. 1989. Folk Psychology and the Explanation of Human Behavior. In Churchland, *A Neurocomputational Perspective*. Cambridge, MA: MIT Press. Pp. 111-128.

Cummins, R. 1983. *The Nature of Psychological Explanation*. Cambridge, MA: MIT Press.

D'Andrade, R. 1987. A Folk Model of the Mind. In Holland, D., and Quinn, N. (eds.), *Cultural Models in Language and Thought*. Cambridge: Cambridge University Press.

Dennett, D. 1969. *Content and Consciousness*. London: Routledge and Kegan Paul.

Dennett, D. 1978a. Artificial Intelligence as Philosophy and Psychology. In Dennett, *Brainstorms*. Cambridge, MA: MIT Press. Pp. 109-126.

Dennett, D. 1978b. Two Approaches to Mental Images. In Dennett, *Brainstorms*. Cambridge, MA: MIT Press. Pp. 174-189.

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

Fodor, J. 1968: The Appeal to Tacit Knowledge in Psychological Explanation. *Journal of Philosophy*, 65, 627-640. Reprinted in Fodor 1981.

Fodor, J. 1975. *The Language of Thought*. New York: Thomas Crowell.

Fodor, J. 1981: *Representations*. Cambridge, MA: MIT Press.

Fodor, J., Bever, T. and Garrett, M. 1974. *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. New York: McGraw-Hill.

Goldman, A. 1989: Interpretation Psychologized. *Mind and Language*, 4, 161-185.

Gordon, R. 1986: Folk Psychology as Simulation. *Mind and Language*, 1, 158-171.

Gordon, R. unpublished, a: Fodor's Intentional Realism and the Simulation Theory. MS dated 2/90.

Gordon, R. unpublished, b: Simulation and the Theory-Theory. MS dated 1/25/91. Presented at Pacific APA, 1991.

Gregory, R. 1970. *The Intelligent Eye*. New York: McGraw-Hill.

Greeno, J. 1983. Conceptual Entities. In Gentner, D., and Stevens, A. (eds.) *Mental Models*. Hillsdale, NJ: Erlbaum. Pp. 227-252.

Hayes, P. 1985. The Second Naive Physics Manifesto. In Hobbs, J., and Moore, R. (eds.), *Formal Theories of the Commonsense World*. Norwood, NJ: Ablex. Pp. 1-36.

Heal, J. 1986: Replication and Functionalism. In J. Butterfield (ed.), *Language, Mind and Logic*. Cambridge: Cambridge University Press. Pp. 135-150.

Johnson-Laird, P. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*. Cambridge, MA: Harvard University Press.

Kahneman, D. and Tversky, A. 1982. The Simulation Heuristic. In Kahneman, D., Slovic, P. and Tversky, A. (eds.) *Judgment Under Uncertainty*. Cambridge: Cambridge University Press.

Kosslyn, S. 1981. The Medium and the Message in Mental Imagery: A Theory. In Block, N. (ed.), *Imagery*. Cambridge, MA: MIT Press. Pp. 207-244.

Leslie, A. 1987. Pretense and Representation: The Origins of "Theory of Mind". *Psychological Review*, 94, 412-426.

Leslie, A. 1988. Some Implications of Pretense for Mechanisms Underlying the Child's Theory of Mind. In Astington, J., Harris, P. and Olson, D. (eds.), *Developing Theories of Mind*. Cambridge: Cambridge University Press. Pp. 19-46.

Lycan, W. 1981. Form, Function and Feel. *The Journal of Philosophy*, 78, 24-50.

Lycan, W. 1988. Toward a Homuncular Theory of Believing. In Lycan, *Judgement and Justification*. Cambridge: Cambridge University Press.

Marr, D. 1982. *Vision*. San Francisco: Freeman.

McCloskey, M. 1983. Naive Theories of Motion. In Gentner, D. and Stevens, A. (eds.), *Mental Models*. Hillsdale, NJ: Erlbaum.

Montgomery, R. 1987: Psychologism, Folk Psychology and One's Own Case. *Journal for the Theory of Social Behavior*, 17, 195-218.

Newell, A. 1980. Physical Symbol Systems. *Cognitive Science*, 4, 135-183.

Nisbett, R. and Ross, L. 1980. *Human Inference*. Englewood Cliffs, NJ: Prentice-Hall.

Olson, D., Astington, J. and Harris, P. 1988. Introduction. In Astington, J., Harris, P., and Olson, D. (eds.) 1988. *Developing Theories of Mind*. Cambridge: Cambridge University Press.

Perner, J., Leekam, S. and Wimmer, H. 1987. Three-year-olds' Difficulty with False Belief: The Case for a Conceptual Deficit. *British Journal of Developmental Psychology*, 5, 125-137.

Pinker, S. 1989. *Learnability and Cognition*. Cambridge, MA: MIT Press.

Pylyshyn, Z. 1981. The Imagery Debate: Analog Media versus Tacit Knowledge. In Block, N. (ed.), *Imagery*. Cambridge, MA: MIT Press. Pp. 151-206.

Pylyshyn, Z. 1984. *Computation and Cognition*. Cambridge, MA: MIT Press.

Ramsey, W., Stich, S. and Garon, J. 1990. Connectionism, Eliminativism and the Future of Folk Psychology. *Philosophical Perspectives*, 4, 499-533.

Ripstein, A. 1987: Explanation and Empathy. *Review of Metaphysics*, 40, 465-482.

Rock, I. 1983. *The Logic of Perception*. Cambridge, MA: MIT Press.

Rumelhart, D., Smolensky, P., McClelland, J., and Hinton, G. 1986. Schemata and Sequential Thought Processes in PDP Models. In McClelland, J., Rumelhart, D., and the PDP Research Group, *Parallel Distributed Processing*, Vol. 2. Cambridge, MA: MIT Press.

Sellars, W. 1963. Empiricism and the Philosophy of Mind. In Sellars, *Science, Perception and Reality*. London: Routledge and Kegan Paul. Pp. 127-196.

Stich, S. 1978. Beliefs and Subdoxastic States. *Philosophy of Science*, 45, 499-518.

Wimmer, H., Hogrefe, J. and Sodian, B. 1988. A Second Stage in Children's Conception of Mental Life: Understanding Informational Access as Origins of Knowledge and Belief. In Astington, J., Harris, P. and Olson, D. (eds.), *Developing Theories of Mind*. Cambridge: Cambridge University Press. Pp. 173-192.

NOTES

1. We are grateful to Professor Gordon for providing us with copies of his unpublished papers, and for allowing us to quote from them at some length.

2. The wind tunnel analogy is suggested by Ripstein (1987 p. 475 ff). Gordon also mentions the analogy (unpublished, b, p. 8) but he puts it to a rather different use.

3. The burglar in the basement example is borrowed from Gordon 1986, p. 161.

4. The evidence Gordon cites includes the tendency to mimic other people's facial expressions and overt bodily movements, and the tendency in both humans and other animals to direct one's eyes to the target of a conspecific's gaze.

5. Gordon 1986, p. 163 ff.

6. Ripstein's account of role of simulation in intentional explanation is quite similar.

> I wish to defend the claim that imagining what it would be like to be in "someone else's shoes" can serve to explain that person's actions.... I shall argue that imagining oneself in someone else's situation ... allows actions to be explained without recourse to a theory of human behavior. [Ripstein 1987, p. 465]

> [T]he same sort of modeling [that engineers use when they study bridges in wind tunnels] is important to commonsense psychology. I can use my personality to model yours by "trying on" various combinations of beliefs, desires and character traits. In following an explanation of what you do, I use my personality to determine that the factors mentioned would produce the result in question.... I do not need to know how you work because I can rely on the fact that I work in a similar way. My model ... underwrites the explanation by demonstrating that particular beliefs and character traits would lead to particular actions under normal circumstances. [Ripstein 1987, pp. 476-477]

7. As Jerry Fodor has pointed out to us, the logical geography is actually a bit more complex than Figure 2 suggests. To see the point, consider the box labeled "Decision Making (Practical Reasoning) System" in Figure 1. Gordon and Goldman tell us relatively little about the contents of this box. They provide no account of how the Practical Reasoning System goes about the job of producing decisions from beliefs and desires. However, there are some theorists - Fodor assures us that he is one - who believe that the Practical Reasoning System goes about its business by exploiting an internally represented decision theory. If this is right, then we exploit a tacit theory each time we make a decision based on our beliefs and desires. But now if we make predictions about other people's behavior by taking our own Practical Reasoning System off-

line, then we also exploit a tacit theory when we make these predictions. Thus, contrary to the suggestion in Figure 2, off-line simulation processes and processes exploiting an internally represented theory are not mutually exclusive, since some off-line simulation processes may also exploit a tacit theory.

In the pages that follow, we propose to be as accommodating as possible to our opponents and to make things as hard as possible for ourselves. It is our contention that prediction, explanation and interpretation of the sorts we have discussed do not use an off-line simulation process, *period*. So if it turns out that Fodor is right (because the Practical Reasoning System embodies an internally represented theory) *and* that Gordon and Goldman are right (because we predict and explain by taking this system off-line), then we lose, and they win. Also, of course, if Fodor is wrong about how the Practical Reasoning System works but Gordon and Goldman are right about prediction, explanation and interpretation, again we lose and they win. So as we construe the controversy, it pits those who advocate any version of the off-line simulation account against those who think that prediction, explanation and interpretation are subserved by a tacit theory *stored somewhere other than in the Practical Reasoning System*. But do keep in mind that we interpret "theory" broadly. So, for example, if it turns out that there is some non-sentence-like, non-rule-based module which stores the information that is essential to folk psychological prediction and explanation, and if this module is not used at all in ordinary "on-line" practical reasoning and decision making, then we win and they lose.

It might be protested that in drawing the battle lines as we propose to draw them, we are conceding to the opposition a position that they never intended to occupy. As we have already noted, Gordon and Goldman expend a fair amount of effort arguing that a tacit theory is not exploited in folk psychological prediction and explanation. Since they think that the Practical Reasoning System *is* exploited in folk psychological prediction and explanation, presumably they would deny that the Practical Reasoning System uses an internally represented decision theory. So it is a bit odd to say that *they* win if Fodor is right about the Practical Reasoning System and they are right about off-line simulation. This is a point we happily concede. It is a bit odd to draw the battle lines in this way. But in doing so, we are only making things more difficult for ourselves. For we must argue that *however the Practical Reasoning System works* we do not predict and explain other people's behavior by taking the system off-line.

8. Gordon (unpublished, a), Sec. 3.5.

9. Actually, the developmental facts are rather more complicated than Gordon suggests. For, as Leslie (1988) emphasizes, children are typically able to appreciate and engage in pretend play by the time they are two and a half years old - long before they can handle questions about Maxi and his false beliefs. It is not at all clear how the off-line simulation theory can explain both the early appearance of the ability to pretend and the relatively late appearance of the ability to predict the behavior of people whose beliefs and desires differ from one's own.

10. It's worth noting that both Gordon's account and Leslie's "predict that only those children who can engage in pretend play can master the concept of belief." [Gordon 1986, p. 169] This may prove a troublesome implication for both theorists, however. For it is not the case that *all* autistic children do poorly on the Wimmer-Perner test. In the original study reported by Baron-Cohen, Leslie and Frith (1985), only 16 out of 20 autistic subjects failed the Wimmer-Perner test. The other 4 answered correctly. The investigators predicted that these children "would also show evidence of an ability to pretend play." (p. 43) Unfortunately, no data was reported on the pretend play ability of these subjects. If it should turn out that some autistic children do well on the Wimmer-Perner test *and* lack the ability for pretend play, both Gordon's explanation and Leslie's would be in trouble. If the facts do turn out this way, advocates of the theory-theory will have a variety of other explanations available. But it is much less clear that the off-line simulation account could explain the data, if some autistic children can't pretend but can predict the behavior of people with false beliefs.

11. Another experiment reported by Perner et. al. (1987) provides some additional evidence for this conclusion. In the first part of the experiment children were shown a box of Smarties (a type of candy), and asked what they thought was in the box. All of them answered that the box contained Smarties. They were then shown that the box contained a pencil, and no Smarties. After this the children were asked three questions:

i) What is in the box.

ii) What did you think was in the box when you first saw it.

iii) What would a friend, waiting outside, think was in the box if he saw it as it is now.

Most of the younger children answered (iii) incorrectly; they failed to predict their friend's false belief. But more than half of those who got (iii) wrong answered (ii) correctly. They were able to tell the experimenter that they had thought the box contained Smarties, and that they were wrong. In commenting on this experiment, Leslie notes that

> [d]espite the ability to *report* their false belief, these 3-year-olds could not understand where that false belief had come from.... Despite the fact that they themselves had just undergone the process of getting that false belief, the children were quite unable to understand and reconstruct that process, and thus unable, minutes later, to predict what would happen to their friend. [Leslie 1988, pp. 33-4.]

12. Pylyshyn 1981 & 1984. It is perhaps worth noting that we are using the term "cognitively penetrable" a bit more loosely than Pylyshyn does. But in the present context the difference is not important.

13. We are grateful to Jerry Fodor for his helpful comments on an earlier version of this paper. Thanks are also due to Joseph Franchi for help in preparing the Figures.