The Status Of Simulation Theory in The Interpretation Debate

Joanne Downs

A Thesis

In

The Department

Of

Philosophy

Presented in Partial Fulfilment of the Requirements

For the Master of Arts Degree at

Concordia University

Montreal, Quebec, Canada

October 1998

0-612-39054-3

ABSTRACT

The Status of Simulation Theory in The Interpretation Debate

Joanne Downs

There is a debate going on in the contemporary philosophical literature concerned with our folk-psychological capacities, that is, how we explain, predict and interpret the behavior of others, and how we ascribe mental states such as beliefs and desires to each other. The purpose of this thesis is to examine the status of the two opponents of this debate, simulation theory and theory-theory, concerning the plausibility of their explanation of self knowledge, our capacity to attribute mental states to ourselves and to each other. A preliminary concern is to determine whether or not simulation theory should be considered a plausible rival to theory-theory, the presently dominant theory of our folk psychological capacities.

Dedication

I would like to thank Dr. Murray Clarke for having given me the idea to write on the Interpretation Debate topic in the first place, and Dr. Bernier for being my supervisor. It was his constant demonstration of moral support and conscientious editing style that enabled me to produce the present document.

I would also like to thank Bill Massicotte and Brian MacPherson, for convincing me many years ago that I had something to contribute to the field of philosophy, and for their continuing support in my academic endeavours.

I thank the faculty and staff of the philosophy department and especially Eudene, for rescuing me from many potentially hazardous bureaucratic situations.

This thesis is dedicated to the Downs's, and to my technical support, Willy 'Wolfboy' Morelli, the unbeliever.

Table of Contents

INTRODUCTION

How is it that ordinary folk are able to predict and explain each other's behavior in a reasonably reliable and adequate way? What allows us to be able to attribute mental states to others? What sorts of cognitive capacities are required by our 'folk psychological practices' ? By this expression I mean the ability we folk have of predicting and interpreting the behavior of others, and of attributing beliefs to others. Also included in the folk psychological practices are the ability to identify one's past, current and future mental states, the ability to predict one's own behavior, and the ability to engage in conditional planning, a mental exercise whereby one imagines different future scenarios in planning an activity or an outing. From now on, I will be using the shorthand term 'folk psychological practices' to mean all of those practices in which we use the abilities mentioned above. The endeavor to develop a theory that would account for these folk psychological practices has recently become the topic of a debate in the philosophical literature.

Until recently, the dominant viewpoint has been that our folk psychological practices depend on some sort of theory. For instance, in order to determine why my friend refused the offer of something to eat the other night at a party, I might refer to my folk psychological theory, which contains the platitude: people refuse food when they are not hungry. I can then surmise that perhaps my friend has refused the offer because she is not hungry. Variations on this 'theory' theme are numerous: the theory may or may not be accessible to consciousness, it might either be innately acquired or culturally learned, it might resemble a theory in its dynamic and static features, or it might only be considered a theory if the word 'theory' is taken in its widest sense.[1] There are several historical sources that have contributed to the emergence of the idea that our folk psychological practices depend on a theory. David Lewis, in an influential article, attempted to answer the question: how is meaning given to mental terms such as 'belief' and 'desire' ? (Lewis,1972). According to his account, theoretical terms get their meaning from being embedded in a theory. He suggested that if we view mental terms such as 'belief' and 'desire' as theoretical terms, the idea that theoretical terms get their meaning from being part of a theory can then be extended to the idea that mental terms could get their meaning from being part of a folk psychological theory which implicitly underlies our folk psychological practices. In addition to Lewis' view, the question of how a theory of mind might be psychologically involved in our folk psychological practices was addressed by Jerry Fodor (Fodor,1968). He suggested that our 'folk theory' is a set of law-like generalizations, organized in sentence-like form and somehow represented in the mind, and predominantly innate. One possible challenge would be to require that proponents of the view explicitly state the generalizations.[2] A way to circumvent this challenge is to claim that the folk psychological theory is largely tacit, and thus unavailable to consciousness. Fodor has drawn an analogy between the folk psychological theory that we use in our folk psychological practices and the linguistic theory that we use in the understanding of language. The point of the analogy is that if we judge the grammaticality of a sentence by deploying a tacit theory, then perhaps we also deploy a tacit psychological theory to interpret the behavior of others. That is to say, we are not aware of possessing and using such a theory. Moreover, in other areas of cognitive science it is also speculated that the folk have and deploy a 'folk physics', in order to explain common physical phenomena, for example, how heat escapes from a room through poorly insulated windows. The above remarks are a brief sample of the theories put forth as attempts to explain our folk psychological practices. These attempts are similar in that these practices are said to depend on some sort of theory.

Increasing dissatisfaction with the idea that our folk psychological practices depend on some sort of theory, among other factors, has led to the emergence of an alternative theory, known as simulation theory. Advocates of this new alternative, which began to emerge around 1986, are not persuaded that our folk psychological practices are represented in the form of a theory, or even that these practices depend on an underlying theory.[3] This rival hypothesis is a practical alternative in the sense that in interpreting another's behavior, for example, what I do is put myself 'in the shoes of' the other person, and decide what I would do in those circumstances. It could be considered a plausible rival in the sense that on this view I would not need to make reference to, nor deploy, any theory.

Discussion between proponents of these two alternatives has recently evolved into the form of a debate. The traditionally dominant explanation, the theory subsumption explanation, is usually called the 'theory-theory'. The common claim among philosophers on this side of the debate is that the folk deploy some form of a theory in order to interpret others. The recently developed rival account, simulation theory, claims that in order to interpret the behavior of another we rather undergo a process of simulation, where we mimic the context of the other person by trying to track their mental state sequence. The process has been characterized by Gordon as "...Practical simulation imitates real life." (Gordon,1986:62).

Certain advocates of the theory-theory side have drawn attention to some possible prima facie problems in the simulation account, which will be discussed in chapter one. The prima facie theory-theory objection has to do with the fact that most simulation accounts (except Gordon's) are based on an analogical inference. That is , the simulator imagines himself in the target's shoes, identifies his own mental states then infers the mental states of the target based on an analogy from his own case. The theory-theory objection is that in order to identify his own mental state the simulator must already possess the theoretical concepts 'belief' and 'desire', for instance, of which the mental states are instances of. Possession of the theoretical concepts, it is argued, requires a theory. The question is then: Does the simulation side need to postulate concept mastery, as well as a theory at source, in order to get off the ground? This question will be the central point of discussion in chapter one of the thesis, and I will argue that simulation theory can claim concept mastery without accepting the further implication of a background theory, and still be able to stand on its own. My general aim in the thesis is to show that simulation theory is a strong contender in the Interpretation Debate, and also to defend a particular version of simulation as the most plausible, namely that of Alvin Goldman. I will not be considering the recently developed 'hybrid' theories that are a mix of simulation and theory-theory, since it would be beyond the scope of the present discussion. However, further research into the debate would warrant such a consideration, in order to examine the plausibility of a mix of the two theories.

My thesis will proceed along the following lines. In chapter one I present a description of three versions of the new rival, simulation theory, in order to compare and contrast the three accounts. To that end, I use the main tenets of Alvin Goldman's version (1989, 1992a, 1992d, 1993b), as well as the developmental account of a psychologist, Paul Harris (1992, 1995, 1996). The chapter ends with a discussion of Robert Gordon's account (1986a, 1995, 1996). In addition to the description of these three accounts, I also underline the motivations that led each of these three authors to initially develop a rival to the theory-theory. Finally, in the rest of the chapter I argue that simulation theory is not just another version of theory-theory, and that it should be considered as a plausible contender in the debate. I also argue that we need not endorse Gordon's strong version in order to get out from under the above theory-theorist objection.

There is much less agreement among theory-theorists on questions of the nature of the theory-theory, how it is acquired, and the like. In chapter two, I present two versions of the theory-theory that are considerably different from each other. I begin with that of Alison Gopnik (Gopnik,1996; Gopnik and Wellman,1992), which is considered the strongest version, and then I discuss the general version of Stich and Nichols. My strategy is to show just how different various versions of the theory-theory can be, particularly in the construal of the notion 'theory' involved in their respective views. The chapter revolves around this issue of 'theory', specifically how it is defined by these authors, with the goal of showing that the theory-theory claim that simulation is just another of its versions does not hold much weight. The primary reason is the lack of consensus on a reasonably restricted sense of the term 'theory'.

In chapter three, my aim is to submit both simulation and theory-theory to the test. The test here is an issue of relative importance to an account of our folk psychological practices, since it is the issue of how we go about ascribing mental states to ourselves and others. My strategy in this chapter is to discuss how Gopnik from the theory-theory side and Gordon and Goldman/Harris from the simulation side explain mental state ascription, with the aim of determining which side gives the better account. It turns out that the accounts of Gopnik and Gordon are open to serious objections and that the Goldman/Harris account is the most plausible of those examined.

ENDNOTES

[1] The issue concerning how the notion of 'theory' is construed by each theory-theorist is discussed in chapter two.
[2] See for example Goldman, 1989.
[3] See, for example, Jane Heal,1986; Robert Gordon,1986; and Alvin Goldman,1992.

CHAPTER ONE

CAN SIMULATION THEORY STAND ON ITS OWN?

As was pointed out in the introduction, in the mid 1980's a new alternative to the theory-theory arose, also concerned with how we are able to explain, interpret and predict the behavior of others. The main focus of this first chapter is to determine whether or not simulation theory is really an alternative, and thus a rival, or whether it is just another version, albeit a special version, of the theory-theory.

One might wonder why such a question should arise, since there is a debate going on in the contemporary literature concerning precisely these two sets of theories, and the question in this debate is: which side gives the better account of our range of folk psychological practices? Here it would appear that simulation theory has already established itself as a genuine, plausible rival to the theory-theory. Why should it then have to submit itself to the preliminary consideration of being a plausible rival? The answer to this question stems largely from the fact that the two sides are on uneven territory, so to speak. The theory-theory has been around for a good many more years than simulation has, and so is considered to be the *status quo* theory. Simulation theory must prove itself as more than a plausible contender in order to replace, as its defenders wish it to, the theory-theory side.

For this reason, this chapter begins with the new alternative, simulation theory, in particular a characterization of three of its important versions, those of Alvin Goldman, Paul Harris and Robert Gordon. Once I have given the reader an idea of what each author's theory is about, it is then possible to view Harris and Goldman's accounts as similar, with Gordon in a separate category on its own. The reasoning behind this comes from Gordon himself: he wishes to distinguish himself from all other simulationist accounts, claiming that other versions incorporate certain assumptions that he does not wish to use, since he views them as problematic. On Gordon's view, these assumptions are problematic because he thinks that accepting them would force simulation theory to require some type of theory. These assumptions are the focus of the latter part of the chapter, the middle part is concerned with drawing a sharper contrast between the accounts of Goldman and Harris on the one hand, and Gordon on the other.

The onus behind the endeavor of determining whether or not simulation theory is a plausible rival to the theory-theory comes from none other than theory-theory "loyalists" (to a certain extent). If simulation theory wants to eventually replace its rival, it must obviously answer to certain criticisms from the present dominant theory. The main criticism coming from the dominant side happens to be that there is no distinct new rival. How much of this criticism (all centered around the objection that simulation must depend on some kind of theory or theoretical constructs) is fear of being replaced and how much is true shortcoming on the part of simulation theory will be determined by the end of this chapter.

The general idea behind the simulation theory is that we predict or interpret the behavior of others by metaphorically putting ourselves in their shoes. There is by no means one single version of the simulation theory that all of its advocates endorse. On the contrary, and as will be seen in this chapter, there are quite striking contrasts, especially between that of Gordon and those of Harris and Goldman. However, there is one common claim among all simulation accounts, and that is that our folk psychological practices can be explained without having recourse to a theory. This common claim is what has distinguished simulation theory from theory-theory in the first place, and allowed it to become a contender in the present debate in the philosophical literature. This common claim, it should be noted, is central to the dispute of whether simulation theory is just another version of theory-theory, which it would be if it had to depend on a theory.

There is also a general assumption shared by all versions of simulation theory, namely that others are psychologically similar to ourselves. Without this crucial assumption, simulation theory could never get off the ground, for how am I to predict your behavior in a situation based on how I would react in the same situation, if you and I are not psychologically similar? There would be no basis for comparison, and thus no common basis for prediction or interpretation. Goldman makes this assumption explicit, modifying Grandy's humanity principle which states that "the imputed pattern of relations among beliefs, desires and the world be as similar to our own as possible." (Goldman, 1989:81). This assumption has been the focus of an objection by the theory-theory side. They argue that the set of psychological similarities assumed by the simulationist is itself constitutive of a theory, hence the non-theoretical alternative that is supposed to be simulation is not really so. The simulationist does not commit himself to this claim, however, the assumption is merely meant as a general constraint on the process of simulation, and nowhere during that process is there a need to consult this supposed theory. Moreover, is it so unreasonable to take it as a given that we all have certain common cravings and beliefs without assuming that these cravings must come from a theory, just like the idea that other biological species might be similar?

1. Goldman's 'Theory-Driven' and 'Process-Driven' Simulation

The general idea behind Goldman's version of simulation is that it is "an intensively used heuristic, one on which interpretation fundamentally rests." (Goldman,1989:83). His version is often referred to as 'Model-Model' simulation, in the sense that I would use myself as a 'model' of someone else's mental and behavioral life. The idea here is that simulations of the (human) target model are based on inference from an analogy, stemming from using one's own 'would-be' or 'as-if' mental states and behaviors as a model of the target's. As we'll see in section three, this account of things is different from how Gordon puts it, and this is because he does not make use of an analogy from the simulator to the target, nor does he need to postulate an inference based on this analogy.

It is useful at this point to reiterate Goldman's now classic Crane/Tees example to illustrate how Model-Model simulation works (Goldman,1989:83): Mr. Crane and Mr. Tees are both scheduled to catch a plane, and they both arrive late at the airport. Mr. Crane misses his plane by three hours, and Mr. Tees misses his by five minutes. The test question is: who do you think will be more upset? The near perfect consensus (that Mr. Tees would be more upset) of the answers given to this vignette suggests to Goldman, at least, that it is highly unlikely that all of the test subjects had a folk psychological theory that would contain this particular mundane platitude. In this case it would be something along the lines of "to miss your plane by a short period of time is more frustrating than to miss it by a long period of time". He suggests instead that we 'put ourselves in each of the men's shoes', see how we would feel and then infer by analogy how each of the men would feel.

Goldman realizes that not all simulations are done using this method of inference by analogy, for I could not simulate a rock rolling down a hill by inferring its behavior by analogy from my own case. Daniel Dennett has raised this objection using a suspension bridge as an example (Goldman,1989:84-5). He rightfully objects that I could not successfully simulate a suspension bridge without having at least some theoretical knowledge of how suspension bridges work. Taking this point into consideration, Goldman has postulated two types of simulation, one for simulating non-humans, 'theory-driven', and one for humans, 'process-driven'. Dennett would use theory-driven simulation to determine how a suspension bridge would fare in an earth quake and Goldman would use process-driven simulation to determine who would be the more distraught, Mr. Tees or Mr. Crane. Goldman realizes that some human simulations may be theory-driven, as he states it "Inductine or nomological information is not wholly absent, but it is sparser than the folk-theory approach alleges." Goldman,1989:82)

There are two criteria that must be met in order for process-driven simulation to be successful, according to Goldman, both stemming from the humanity principle mentioned earlier. Intuitively, one would think that in order to simulate the behavior of another, that other would have to be reasonably similar to oneself. This is precisely Goldman's first criterion: that the processes driving the two systems, simulator and target, be as similar as possible. If Dennett is simulating a suspension bridge, he will obviously use theory-driven simulation rather than process-driven, since he is not made of material that is reasonably similar to a suspension bridge. Moreover, in attempting to simulate a bridge he is violating the humanity principle, necessary for all human simulations concerned with here, at any rate. The other condition necessary for a successful simulation is that the two systems be in similar initial states. This seems reasonable particularly if I am going to be attributing mental states to someone during the process of simulating them. I cannot hope to achieve any degree of accuracy unless I am in a reasonably similar initial state. For instance, I would need to get into the initial state of anger that the target was feeling due to what she thought was an unfair call, if I want to accurately interpret her having just punched a referee in a basketball game. In a nutshell, here is an abridged view concerning how Goldman describes process-driven simulation.

> The initial step is to imagine being "in the shoes" of the agent, e.g., in the situation of Tees or Crane. This means pretending to have the same initial desires, beliefs, or other mental states that

> the attributer's background information suggests the agent has. The next step is to feed these pretend states into some inferential mechanism, or other cognitive mechanism, and allow that mechanism to generate further mental states as outputs by its normal operating procedure. ...the output state should be viewed as a pretend or surrogate state... Finally, upon noting this output, one ascribes to the agent an occurrence of this output state. (Goldman,1992a:189).

With regard to the concern of how the simulator is able to achieve an initial state of mind that is reasonably similar to the target's, again the humanity principle is relevant. If two human beings can be considered reasonably psychologically similar, then is it a further leap of faith to infer that they would share common cravings, beliefs and desires, especially if these desires are considered given that the two are in a similar environment? A good example of a commonality among most humans is the restaurant script. Most people have been in a restaurant at least once in their lives and thus know the ordering of events, such as waiting to be seated, receiving the menu, ordering the meal, and so on. Thus if I were attempting to simulate a target in a situation occurring in a restaurant, I could probably be reasonably accurate in my interpretation just because I know what goes on in the restaurant situation.

Goldman's version of simulation relies heavily on two concepts: pretence and empathy. As can be seen from his three-step description of the process, not only is pretence important in the construction of the initial state of the target, but also the entire offline stage of the process can be considered as a pretend episode, where feigned environmental and situational conditions are considered by the simulator, in order to generate would-be behavioral output that is a model of the target's behavior. As it turns out, the cognitive system must be taken offline for most simulation episodes, the only time that it remains online is when a decision has been made to execute a particular plan within the multitude that are considered during the act of conditional planning. All other simulations, concerning the prediction and explanation of behavior, or the attribution of a particular mental state to another, require that the system be offline, for the obvious reason that these simulations are interpretations and explanations of the target and not actual behavior The important role of pretence in simulation could be considered as a major strength of Goldman's theory, since it would account for the relevance of pretend play in children. In fact, as will be seen in Harris' account, this is the main claim of Harris: that simulation is just a more mature and sophisticated capacity that has developed from pretend play in childhood.

Empathy, which is also central to Goldman's account, can be considered as a strength since it plays a major role in the simulating of the emotional states of others. Before this strength can be examined, however, it is necessary to understand how Goldman uses the term. The fact that Goldman's entire theory of simulation depends on the concept of empathy may seem initially strange, especially when one considers that the concept is so vaguely defined and ill understood as to be considered meaningless. Given the vagueness of the term, it might be argued that one cannot employ the term in one's theory without at least defining it. Goldman is aware of this problem and has attempted to define the term. He uses it in two commonly construed senses as he sees it (namely in a broad sense as well as a narrow sense of the term). He construes empathy in the broad sense when he claims that it underlies the entire simulation process. Here 'empathy' is meant to connote the commiseration with the two men that I would engage in when I am 'putting myself in Mr. Tees or Mr. Crane's shoes' for example. Construed in this broadly defined way, it is just the act of attempting to denote the process of simulation. In the narrow sense of the term, 'empathy' is meant to connote an emotional comradery of sorts, as I would engage in if I were to simulate the reaction of someone who had just been told that his parents have been killed in a car accident. Here is how Goldman differentiates the two meanings:

> Until now I have used the term "empathy" to denote the process

> of simulation. But "empathy" typically has a narrower meaning, one specifically concerned with affective or emotional states. To empathize with someone, in its most frequent sense, is to sympathize or commiserate, which involves shared attitudes, sentiments, or emotions (Goldman,1992a:196).

Having this doubly important role in the simulation process, as the underlying mechanism and as the dominant mode in the interpretation of emotions, empathy ends up being a major strength of Goldman's theory, for it serves to account for how we are able to understand, predict and interpret the emotions of others. In contrast to the theory-theorist's account, it seems much more intuitively plausible to say that we empathize with the emotional state of another than to say that we search through the files of our folk psychological theory for the appropriate emotion that we theorize the target to be feeling at that moment. If the empathetic mode thus turns out to be a better account of how we simulate emotions in others than the theory-theory explanation, then the simulation side gains more and more plausibility as a result of this, and there is less reason to accept the claim that simulation theory is just another version of the theory-theory, for it offers a reasonably plausible alternative to the theory-theory account.

2. Harris and the Development of the Simulation Heuristic

In my differentiation of accounts within the category of simulation theory I have placed Harris in a category along with Goldman, and Gordon in a self-imposed category of his own. This is because Harris' account is also of the Model-Model type, where an inference must be made from the simulator's case to the target's. Harris and Goldman's versions complement each other in an interesting way in that Harris plots the developmental course that the simulation capacity takes, and Goldman's continues where Harris leaves off, into the matured capacity that is used in adulthood.

Harris sees simulation as a heuristic that becomes increasingly well honed during the course of the child's development due to incremental increases in imaginative flexibility (Harris,1992:216). These incremental increases take the form of a set of four chronological age-specific stages, after which, at the age of approximately five years, the child should have a nearly fully matured ability to simulate. Here follows an outline of the four stages: In stage one the child is able to *echo* another's perceptual perspective toward an object. That is to say, her gaze will be directed toward an object that she notices her mother is looking at. The mechanism behind this ability is assumed to be a special in-built mechanism that facilitates the occurrence of the "shared gaze" phenomenon, where a baby's attention and perception is directed to the object of her mother's gaze. In the second stage, the shared gaze phenomenon is enhanced so that the child can now differentiate her own perceptual perspective from that of another person. Here there seems to occur a separating off of the child's intentional stance from that of others, or to put it another way, the child now realizes that others might have a different visual perspective from her own. It is here that the empathy mode discussed by Goldman presumably starts up, for the child is now able to understand that other perspectives exist but it is not until the third stage, where she is able to imagine a situation that is not in the current visual field, that she can experience fully fledged empathy. It is this newfound ability to imagine a person that is not herself, in a situation that is not currently the case (in other words, an imagined situation), which allows the empathy mode to be fully activated.

Stages three and four, the development of imaginative flexibility, are a natural extension of the child's tendency to pretend play. They are an elaboration of the previous stages in that they entail the ability to imagine someone else's perspective, and then further the ability to imagine an intentional stance toward a counterfactual situation. In stage three the pretence capacity begins to play a role, for instance in the ability to imagine an object, or imagine the father's perspective toward an object that is not currently in the visual field. Step four brings about an increase in imaginative power or flexibility, evidenced by the child's ability to entertain an imagined situation that runs counter to the current one.

The relevance of taking the cognitive system offline begins at stage two, whereas up until that point it could be said that the child operates with the system exclusively online, since she has not yet acquired the understanding that there exist perspectives that are separate from her own. From that point on, however, offline simulation is certainly the dominant mode, considering that it is mostly used for the prediction, explanation, and interpretation of other people's behavior rather than one's own.

Given this developmental sequence, it now becomes easy to see both how and where some of Goldman's elements have developed. The empathetic mode arises as soon as the splitting off has occurred in the child between his or her own perspective and that of another person, and has fully matured once the child is able to imagine the point of view of another. Once the child has acquired this ability, it seems a small further step to be able to take the cognitive system offline, acquired in stage two, and commiserate with the perspective of another person.

It can also be seen from Harris' four stages how Goldman's Model-Model simulation develops in the child. As was discussed earlier, on Goldman's account the simulator makes an inference as to the target's behavior based on an analogy drawn from his own case to the target. The source of this tendency to use oneself as a model might come from stage three, when there is an increase in imaginative flexibility coupled with the development of the empathetic mode. With these two elements in place, it becomes possible for the child to imagine the situation of another person, and empathize, particularly with the emotional situation of another person.

Both simulation theory as well as theory-theory make clearly defined different predictions concerning the development of mental state attribution in the child. Although it is less often discussed in the literature, included within the set of folk psychological practices that both theories endeavor to account for are those of mental state attribution, both to oneself and to another person. Simulation theory, at least on the accounts of Goldman and Harris, requires this mechanism for it is necessary in the prediction, explanation and interpretation of the target's behavior. For instance, if I were trying to interpret the behavior of someone who was standing within two feet of a poisonous snake, I would need to identify my own mental state, presumably fear, in order to infer correctly that the target also experiences fear, and this is why he has bolted from the scene. Theory-theory, while it does not necessarily require mental state self-identification because it does not postulate the inference mechanism from 'me to you' (simulator to target), nonetheless must account for the fact that it is necessary to identify mental states in others as part of the process of behavior prediction and interpretation.

According to the versions of Goldman and Harris, mental state attribution to others develops after self-ascription, because the child can only imagine the perspective of another well after she can entertain her own perspective. According to theory-theory, both abilities should arise at the same time, since both cases depend on a theory. (This important difference between simulation and theory-theory will be discussed in more detail in chapter three.)

3. Gordon's "Inference-less" Simulation

Having acquired a general understanding of Goldman and Harris' accounts, it is now useful to discuss Gordon's account with a view toward sharpening the contrast between the two types of simulation accounts, that is, Goldman and Harris on the one hand and Gordon's self-imposed category on the other. With the contrasts in mind, it is then easier to see, first, whether or not Gordon's account makes sense without the elements of the Goldman/Harris version, and second, whether either of the two types are able to escape the charge that it is just another version of the theory-theory.

Generally speaking, Gordon views simulation theory as a worthy opponent to theory-theory in that, as he puts it" ...simulation imitates real life." (Gordon,1995:63). His own version has evolved quite considerably over the years, having originally been introduced under the title "hypothetico-practical reasoning" (Gordon,1986), in order to emphasize its natural extension from our ordinary cognitive and reasoning processes. For instance, he gives the mundane example of predicting his own immediate future behavior, like predicting that he will take a sip of his coffee soon after he has lifted it up to his mouth, when he has just formed the intention to do so. Conditional planning is another natural extension of our ordinary reasoning process, for instance I might imagine all the possible alternative activities I could engage in this weekend before picking just two. It is then a small further step to get from predicting one's own behavior to predicting the behavior of someone else, or from conditional planning to interpreting the behavior of another on this view, for it is suggested that the same underlying process is being deployed in say, conditional planning and behavior interpretation.

.Gordon's earlier version has since evolved into the "radical" form that it is today. Being of such a radical nature, it is often deemed the strongest version of simulation. The most recent version comes in two types, the general idea behind this move is to distinguish simulation in its original form, 'total projection', from simulation with the necessary adjustments, called 'partial projection.'(Gordon,1992) Partial projection, or patching total projection as it is sometimes called, entails making physical or other adjustments in order to close the physical/temporal gaps occurring between the simulator and target, or imagining personality or factual changes in order to better match the target's personality. Total projection, the default mode, is used when one is simulating a target that one is in close spatio-temporal proximity to; in other words, when the simulator is standing right beside the target. Gordon gives the following as an example of total projection. Two people are hiking up a mountain, one friend (the target), ahead of the other on the trail, suddenly turns around and runs back down the trail. The other friend (the simulator) wondering why the first friend has displayed this bizarre behavior subsequently sees the reason for it: there is a bear on the trail just ahead of him. There is no need in this situation for the simulator to make any adjustments, i.e., patch total projection, because total projection is accurate enough to provide an explanation for the target's behavior. Here the slight disparity in physical proximity is not large enough to lessen the accuracy of explanation.

The apparent difference between Gordon's two types of simulation and the Model-Model type of Goldman and Harris is in how each version defines the simulation of another person: for Goldman/Harris, it is to put yourself in their shoes and then make an inference from an analogy based on your own would-be behavior. For Gordon's total projection, it is to simulate the target, but in the shoes of the simulator, (since the 'shoes', i.e., the situation, are the same) and determine directly what his or her behavior is. Of course, Gordon realizes that not all simulations occur when the target is standing right beside the simulator, and so he proposes another mode, 'partial projection'.
Partial projection, discussed in more detail further on, consists in simulating the target in the target's shoes, and also determine directly what the target's behavior might be while

making the appropriate adjustments for differences that occur between the simulator and the target's situation. While I think that Gordon's total projection is basically of a Model-Model type, as I argue in what follows, Gordon insists that there is a difference between the two. He claims that neither of his two types are based on an inference premise, and thus that neither is of the Model-Model variety. In other words, neither type would be based on a model (the target) that is inferred from another model (the simulator). The issue here is whether there really is a difference between Gordon's total projection and the Model-Model version of Goldman/Harris, or whether Gordon is just using word play to feign a difference.

To reiterate, according to Gordon, total projection is the default mode that the simulator will perform, and will only be accurate if no adjustments are required, i.e., when the simulator and target are in the same situation. The second type of simulation, 'partial projection' is that of 'patching' total projection, that is, the act of making further adjustments on the default mode of total projection in order to improve accuracy. The need to patch total projection would include cases where I am not in close proximity to my target. These adjustments take the form of physical/temporal adjustments, institutional role switching, and retrieving relevant knowledge about the target. So that, for example, I would move closer to my target if what is making him react in such a way is beyond my visual field, or adjust my temporal dimension if the need arises. If I am a chemist by profession trying to simulate an artist, I would have to switch institutional roles, suppress my urge to be exacting and precise, and instead let my creative ability flourish. If what is making my target react in a particular way has to do with knowledge about her, such as the fact that she is afraid of dogs, then I would have to retrieve this fact in explaining why she has crossed the street against a light, in heavy traffic, when the place she wants to go is on the side of the street she is already on.

One might wonder why Gordon has felt the need to postulate all these intricate adjustments instead of going with total projection by itself. The reason might have something to do with the fact that Gordon's version of projection must rely heavily on environmental cues, overt behavioral cues, and institutional roles, since the simulator has no recourse to the target's inner states, for these would have to be gotten from an inference from an analogy through introspective access to the simulator's mental states, and Gordon's version is not this type of Model-Model simulation. Since his version relies so heavily on external cues, he needs all these adjustments in order to get the accuracy that would otherwise be obtained through the simple claim of inference from analogy.

It could also be the case that Gordon has postulated the further act of patching total projection for a reason beyond the fact that it improves accuracy. It is possible that he has included the second type of projection to further sharpen the contrast between his and all other simulation accounts, i.e., in order to make explicit that he does not need inference from analogy, along with all of the implications that this claim has.

Assuming, for the moment, that there is a difference between Gordon's total projection and other Model-Model types, the first marked contrast between the Gordon camp and the Goldman/Harris camp is this: a main component of Model-Model versions, the inference as to the target's behavior made from an analogy between the simulator and the target, is not one of the claims included in either of Gordon's two types of simulation. Gordon does not wish to have this claim as part of his theory because he finds it very controversial. In particular it is the assumption which must be implicitly accepted in order to be able to make the inference in the first place that he wants to avoid. For in order to make an inference as to the target's behavior or brain state based on what my, the simulator's behavior or mental state is, I would have to be able to identify my own mental states, and it is this mental state self-attribution claim that Gordon wants to avoid making, since it would require him to accept two more assumptions that he finds controversial. The first is the direct introspective access assumption that mental state attribution in oneself requires, and the second is the concept mastery of psychological terms assumption.

It is my opinion here that Gordon has painted himself into a corner. One thing is clear, and that is that I can never have direct access to the mental states of another, and so any attribution of mental states to another person must be done by inference. Gordon has left the simulator with no way of inferring the state of mind of the target, since he denies introspective access to the simulator. Merely stating that the simulator simulates the target in the target's shoes instead of simulating himself in the target's shoes, does not by fiat provide a natural entry point into the target's state of mind. So we may wonder how the simulator is to determine what the target's mental states are? Moreover, I would argue, contrary to Gordon, that in total projection the only way that the simulator can have access to an explanation of the target's behavior is by making an inference from analogy, requiring that the simulator identify his or her own mental states, in order to get information upon which to base the analogy. This would mean that Gordon's total projection, at least, is indeed a version of Model-Model simulation, or that total projection suffers from the same problem as partial projection: lack of an entry wedge into the target's state of mind.

The second difference between Gordon and the Goldman/Harris camp also pertains to the process of making an inference based on an analogy. Recall that on the Goldman/Harris account, identification of the simulator's own mental states through introspective access is the key component in simulation, for it provides the simulator with the information upon which to base his inferred interpretation of the target's behavior. Gordon does not wish to include the claim that the simulator uses introspective access as providing an analogy to what the target's state of mind might be, because in order to identify one's own mental state, one must possess the concept of such a mental state such as 'belief' or 'desire', and Gordon's concern is that such a concept would have to be theoretical. Theory-theory advocates argue that mastery of concepts requires understanding of a theory. Gordon finds this line of argument embarrassing and while he may or may not be convinced that concept mastery is constitutive of a theory, he would rather avoid the whole controversial question. In other words, Gordon would rather not include an analogical inference in his theory because of the implications, i.e., that it requires concept possession which might in turn require a theory. Accepting even an analogical inference would leave him vulnerable to the theory-theory claim that simulation is just another version of theory-theory.

I am not persuaded that a convincing argument has been made that one must understand some theoretical meaning of terms such as 'belief' or 'desire', whatever that might be, prior to one's being able to employ these concepts. Though Alison Gopnik claims otherwise, it certainly seems a bit implausible to suppose that a child possesses the theoretical concept of 'belief' before she is able to correctly use the word, and it seems further implausible to suppose that the child uses these words without a mastery of the concepts they express, thus without an inkling of what she is saying. There are two problems here: the first is that it is not clear whether using a concept such as 'belief' or 'desire' goes hand in hand with mastery of the concept. In other words, it is possible to imagine a child that uses a concept in her everyday discourse, yet without a mastery of the concept, in the sense that a philosopher would have mastery of the concept. Further, it is not clear whether concept mastery is constitutive of theoryhood. In other words, that to master a concept is to assimilate the theory that it occurs in. Here there is no need to accept the underlying theory-theory claim that concepts are only mastered or understood once their relation to other concepts, in the context of a general theory, is understood. In fact, on this view of things, the child cannot be said to understand any concept at all until all of the relations between concepts have been established, i.e., until all concepts have been encountered and assimilated. This may not happen at all in a person's lifetime. Goldman argues that concept mastery is not necessarily constitutive of theoryhood , and I am inclined to agree with him (Goldman, 1992:196). He also offers another argument as to why the necessity of theoryhood for concept mastery is implausible. He argues that if each new concept assimilated into the theory disturbs the relations between the existing concepts, then each time a new concept is introduced into the theory will require that the relations be reestablished in order to accommodate the new concept (Goldman,1993:23). On this issue, an adoption of Gordon's strong version would thus be unmotivated, since there is no reason to accept concept mastery as being constitutive of a theory.

It turns out that the two differences between the Goldman/Harris camp and Gordon's version of simulation theory discussed above are also arguments as to why simulation theory might be considered as just another version of theory-theory, i.e., the *prima facie* objection mentioned in the introduction. Another way to put this point is that the two differences above end up boiling down to one objection made by the theory-theorist. On the theory-theorist's account, the objection is that simulation theory must claim mastery of theoretical terms such as 'belief' and 'desire', and that this mastery requires a background theory, in order to give coherence and structure and hence meaning to theoretical terms such as 'belief' and 'desire'. This dependence on a theory at some levels collapses it into just another form of theory-theory. It thus appears that Gordon seeks to differentiate his account from other simulation versions precisely because he believes that these other versions are vulnerable to potentially damaging criticism from the theory-theory side, namely that they are vulnerable to becoming versions of the theory-theory. However, as I have argued, this worry seems unwarranted.

The situation thus far is the following: Gordon has developed his own version of simulation that is markedly different from Goldman/Harris because he does not think that simulation should rest on an analogical inference. His motivation for developing a version that is not based on analogical inference is presumably because the analogy aspect would require that the simulator have direct access to his or her mental states. This would commit Gordon to the claim of introspective access, a claim he does not wish to endorse. Furthermore, he does not wish to commit to the claim of concept mastery, another claim that is necessary to the simulation account if it includes an inference 'from me to you'. His most important motivation in denying that simulation theory requires concept mastery, however, is probably that it requires embedding in a theory.

The reality of the situation, as I see it, is that at least one of Gordon's two types of projection, total projection, is a form of 'Model-Model' simulation and thus must rely on an inference from analogy in order to make sense. I would argue that simulation by total projection on Gordon's account cannot proceed without an analogical inference. For there seems to be no way for the simulator to determine the target's mental state aside from environmental cues, which provide only partial information as to the target's state of mind. Partial projection might be adequate without the added step of inference from analogy, but success rests on whether or not the target is able to successfully recenter his egocentric map onto that of the target, if at all. And what does it really mean, after all, to recenter ones egocentric map? One could argue that Gordon is using word play to feign a difference between what he calls recentering one's egocentric map and what is really the use of an analogical inference based on introspective access. He is not actually arguing for what he claims, since he has not given us any reason to believe that to recenter one's egocentric map is to 'become' the target in such a way that one's mental states are actually those of the target, and no longer one's own.

Moreover, I do not think that Gordon needs to go to such lengths in avoiding the claim of introspective access. His attempt to develop and defend a type of simulation (partial projection) that omits the need for introspective access falls flat, for he leaves himself with no way of gaining an entry wedge into the basis for behavior interpretation, which is often by way of mental state self-identification. Relying solely on the environment for clues, his version looks like a form of behaviorism and would be completely inaccurate for the interpretation of a target who is having an emotional reaction, for instance. In any case, Gordon should not be too worried about introspective access, since *prima facie* it seems plausible that we do have direct access to our current mental states. (Although the role of introspective access in the simulation theory will be discussed in more detail in chapter three.)

It would thus appear, after scrutinizing the theories of three simulationists, that there are indeed contrasts to be made among them. There is a particularly sharp contrast to be made between Gordon on the one hand, and Harris and Goldman on the other. What makes one aspect of Gordon's account (partial projection) so different from the other camp is mainly his attempt to avoid making claims that, on his view, render other versions initially problematic. It turns out that the first of these claims, that we have direct introspective access to our mental states, is not so very damaging a claim to endorse, since the issue, originating in Descartes' day, has not been resolved as of yet, and any plausible view would have to account for the fact that we can identify our mental states in a reliable way. Moreover, a definite link has not been established between the idea of using mental state terms and the idea that this use requires a theory. The second and third objections, pertaining to both Gordon and Goldman/Harris, concern the question of whether or not simulation theory can stand on its own. In both cases the argument is that simulation theory must be a special case of theory-theory, because it must depend, at some level, on a theory. Here I have tried to argue the general point that the process of simulation per se does not rely on a theory. That is, nowhere during the act of simulating a target does the simulator need to refer to a theory. These two objections are aimed rather at the cognitive processes that underlie the process of simulation, but that also underlie our day to day thinking. The fact that I might generally need a background theory to house the concept of the word 'belief' or even to house the belief itself, has no bearing on the process of simulation itself, and the fact remains that the process of simulation itself does not require a theory. There is thus no need for Gordon to go to the lengths he does to avoid making any reference to a theory at all because he cannot escape the claim without also sacrificing a large amount of coherence. His version ends up with a rather large theoretical hole: no entry point into the target's state of mind. I don't therefore believe that simulation theory is in danger of becoming just another version of theory-theory, even on the Goldman/Harris version which relies on an analogical inference from the simulator to the target.

CHAPTER TWO

THE THEORY-THEORY

In the last chapter I attempted to show that although it has been claimed by theory-theorists that simulation is just another version of theory-theory, and thus that it is not a worthy contender in the debate over how our folk psychological practices should be explained, there is no compelling reason to suggest that we should accept that simulation is simply a version of theory-theory. As we saw, the term 'theory' is sometimes construed so widely or inclusively that it spills over into cognitive capacities underlying our folk psychological practices. It may indeed be discovered sometime down the road that much of our cognitive processes do depend on a theory, but unless the definition of theory is made more precise by specifying its role in each of these processes, saying that "we use a theory" is like saying nothing at all. For it is not clear where, exactly, the role of 'theory' fits into the explanation of our cognitive processes. The stronger claim could even be made that many of the theory-theorists who defend a version that is broad and inclusive are not defending an identifiable point of view. My first concern, however, is to argue the following: If the claim, made by theory-theorists, that simulation is a version of theory-theory is to hold any weight, then defenders of theory-theory must get clear on the sense of 'theory' that they wish to be understood in their versions.

Not all versions of theory-theory rest on broad construals of the term 'theory', however, and there is lively discussion between theory-theorists on that side of the debate. One version of theory-theory that occurs at one end of the spectrum, and which has the most restrictive construal of 'theory', is the very daring "Child As Scientist" version developed by Alison Gopnik. At the other extreme of possible construals of the term lies the version that Stich and Nichols defend which uses the term 'theory' in a very broad and inclusive sense, similar to the Chomskian model of linguistic competence according to which competent speakers of a natural language are said to use a tacit theory of grammar.

The increasing dissatisfaction with theory-theory, that has led some philosophers to develop the alternative of simulation, has somewhat to do with the vague notion of 'theory' that is understood by some theory-theorists. For it must be clear which sense is meant by the term in order for the claim that simulation theory is just another version of theory-theory to hold any weight. There are at least two senses in which the word 'theory' could be used, both are taken from Webster's dictionary (1994:1387). One is in the sense employed by philosophers of science, as in "... a systematic statement of principles,", and the other is a more loosely understood sense of the term as in "... a speculative idea or plan as to how something might work". On an inclusive notion of 'theory', my 'theory' that the public transportation slows down whenever I leave to go to work would be considered as much a theory as Darwin's theory of evolution. It is obvious here that my 'theory' corresponds to the much looser construal of the term and is not on a par with Darwin's. The problem is well articulated by Simon Blackburn:

> Part of the difficulty is the Protean concept of a theory, often used so that any activity that ends up with a belief counts as forming it by a process of theorizing. (Blackburn, 1995:275).

In order for the theory-theorist's claim to hold any weight, advocates of that perspective must at least agree on and adhere to one and the same construal of the term, and it cannot be the loose sense of the term. One can thus formulate a constraint on the definition of the term 'theory' that is found to be used in the various versions of theory-theory. The constraint is that the definition or understanding of 'theory' must be reasonably restricted, and that a too loosely defined sense of the term cannot be employed because it is not defensible, otherwise any account of our folk psychological practices would turn out to be, trivially, a version of theory-theory on such a loose construal. Hence an argument I will attempt to make in this chapter is that if there are no versions of theory-theory that meet the constraint mentioned above, then theory-theory is not in a position to argue that simulation theory is just another of its versions.

My aim is to show that Stich and Nichols do not meet the above constraint and thus cannot claim in any real way that simulation theory is just another version of their view, and that while Gopnik does meet the constraint, she is ultimately in no better position, since her construal of 'theory' is rather too restricted, for reasons that will be seen. Once simulation theory is able to get out from under the theory-theorist's objection, it is only a further step to seriously undermine the plausibility of theory-theory as a whole.

1. Alison Gopnik: The Child as Scientist

Gopnik defends a version of the theory-theory that is faithful to the classical construal of 'theory' as it occurs in a specific time period in the philosophy of science.[1] Predominantly developmentally concerned, it is an elaboration of how the folk psychological abilities of children are analogous to the static and functional features of scientific theories, and thus that the child's theory changes in much the same way that theories change.

I think that Gopnik's 'Child as Scientist' account is the best example of a version of theory-theory that is defensible, mostly because it employs a concise construal of 'theory'; i.e., not in a loose sense like my theory of public transportation. For her purposes she has borrowed a sense of the term from a particular era in the philosophy of science, and so it would correspond to the first dictionary definition mentioned earlier, on a par with Darwin's theory of evolution, and thus it entirely satisfies the constraint mentioned above. Gopnik may have opted for a narrow construal of the term because she is aware of the general objection made to defenders of theory-theory: that the variations on the 'theory' definition are so broad as to weaken the predictive and explanatory power of the theory. Perhaps indirectly concerned with this problem, Gopnik cites five characteristics that distinguish her use of 'theory' from other uses. Following the classical framework that she uses for defining a theory, the first two features are those of abstractness and coherence, which together give theories the further two features of explanatory and predictive power. The fifth feature of theories is that they should provide interpretations, i.e., coherent explanations, of the data. Plugging these features into the framework and vocabulary of a child's cognitive capacities, she has described the child's theory of mind in the following way:

> All these characteristics of theories ought also apply to children's understanding of mind, if such understandings are theories of mind. That is, such theories should involve appeal to abstract unobservable entities, with coherent relations among them. Theories should invoke characteristic explanations phrased in terms of these abstract entities and laws. They should also lead to characteristic patterns of predictions, including extensions to new types of evidence and false predictions, not just to more empirically accurate predictions. Finally, theories should lead to distinctive interpretations of evidence (Gopnik &Wellman,1992:234).

Gopnik's overall claim is literally that children are little scientists in the classical sense of how scientists do science. As she has stated it, her theory must demonstrate that children's theories of mind show all the five characteristics that are definitive of theories, classically construed. This is a concise and well-defined aim, the question is whether or not she is able to follow through and indeed show that a child's theory of mind contains the same characteristics that theories do.

Her version of theory-theory might be initially intuitively appealing, but as I will show later in the discussion the major drawback is that the claims it makes are perhaps too restricted, and thus Gopnik must follow through by citing evidence that supports her claims. The analogy between scientists and children, who both have a lot of time on their hands to single-mindedly pursue the question of how the world works, might seem plausible because it shows a kind of continuity between the mind of a child and the mind of an adult. It seems right to think that the scientist need not learn a whole new way of thinking in order to pursue science, but rather that the scientific method should come naturally to him because it is what led him to learn about the world and notions such as gravity, the conservation of energy and such, in the first place.

The problem with this picture of things comes with Gopnik's attempt to make it concrete. It is one thing to paint a vague picture of the child throwing all of her toys out of the crib in order to understand how gravity works, but it is another thing entirely to make this picture concrete by claiming that she forms a theory about this activity, for instance, that the child is able to form the theoretical statement "If I start to cry, someone will subsequently come over to where I am". Even assuming that this picture is correct, it raises the further question: at what point does the child start to form her theory of mind, before of after she has learned to speak? This issue has been taken up by Stich and Nichols, in particular with respect to Gopnik's claim that it is 'theories all the way down', that is, children start to acquire fragments of a theory as soon as 42 minutes after birth.

Stich and Nichols raise some serious doubts about the additional assumption that must be made in order for the idea that it is 'theories all the way down' to make any sense. This is the idea that in order to accommodate these complex theoretical statements that are characteristic of a theory, the child must have a rather large degree of developmental precocity, for instance akin to that of a child that is considered a child prodigy and enters university at the age of twelve. The developmental stages that a child goes through generally seem to be chronologically far behind what is necessary for the child to be exploring the world like a young scientist. Stich and Nichols thus argue that it is implausible to suppose that the child actually meets or has the anomalous precocity that 'theories all the way down' would require (Stich and Nichols, forthcoming).

There is also the issue of the particular conception of science that Gopnik has chosen for an analogy with the child's theory of mind. To the question: 'what is the nature of a theory?', there are many possible answers. The fact that Gopnik has chosen the classical conception of theory means that she must at least demonstrate why it is indeed that particular conception that children follow in acquiring a theory of mind rather than another. She must demonstrate, for instance why children have theories of mind similar to Darwin's theory of evolution rather than similar to my speculative 'theory' about public transportation. She must at least recognize that her construal of 'theory' is a contentious claim in that there is a whole branch of discussion in the philosophy of science devoted to the nature of scientific theory and theory change. Her choice to use a construal of theory that is "... as mainstream and middle-of-the-road as possible" (Gopnik and Meltzoff,1996:33) may end up generating more criticism then I suspect she had imagined. To this end, she has landed herself in the exact situation that Gordon has been attempting to avoid in the last chapter, that is, having to defend or justify a preliminary claim or assumption, because it is contentious, before even getting to a defense of the predictions that the theory itself makes.

In addition to the above five 'static' features of theories, there are also, on Gopnik's account, three characteristic steps to the process of theory change, that are known as dynamic features. Theory change itself is ultimately brought about by increasing counter-evidence to the original theory's predictions. The first step is to deny or ignore this counter-evidence, in order to preserve the integrity of the theory. The next step is to invent auxiliary hypotheses to accommodate the counter-evidence, again in an attempt to preserve the original theory. The final step involves theory change itself, creating a new alternative theory, when the overwhelming counter-evidence has rendered the initial theory unwieldy.

Here again Gopnik has made a very sharply defined claim, that children's theories of mind change in the exact same way that theories, classically construed, change. She also follows through on this claim, in the sense that her account of the development of the child's theory of mind exactly mimics that of classically construed theory change. Her strategy, as she puts it, is "...to adopt the detailed descriptions of theories and theory-change in the philosophy of science, translate them into cognitive psychological terms, and see how well they fit the data." (Gopnik,1996:494-5) To this end, she has worked out a chronological developmental account of the child's theory of mind that contains the structural characteristics (both static and functional) of theories. The theory change process is evidenced by the differences that are evident, for instance between the theory of a two year old and that of a five year old, for mere observation seems to point to the fact that a two year-old and a five year-old are not using the very same theory of mind.

On Gopnik's version of the theory-theory, the two year old is claimed to have a markedly different and much less sophisticated theory of mind than the five year old. Gopnik claims that the sequence unfolds in the following way: The two year old has a non-representational desire-perception psychology or theory of mind. It is almost as if the child operates on a direct link with the world, since the interface of representations is lacking. Thus the child can, for instance, predict her own response in a particular situation, but cannot represent and thus predict another person's point of view concerning the same situation. By the age of three, the child's desire-perception psychology has become mediated by representations, perhaps in part due to the onset of language. At about that same age of three years old, there is also an understanding of belief, but not of a representational sort. Thus the child may be using the phrase

"I think" instead of "I believe", giving an indication that she has not yet grasped the fact that individuals can hold beliefs. By the age of five, the desire-perception psychology in the two year old has been replaced by a belief-desire psychology.

On Gopnik's model, the difference between these two theories of mind of the child has partly to do with the acquisition of the ability to form representations, which are believed to underlie the ability to interpret the behavior of another. For instance, in the three year old, since there is no representation yet, the child will be able to understand his or her own beliefs, but not yet be able to understand that others can hold beliefs that are different, or be able to entertain these beliefs of others. The child at this age fails the false-belief task, which is designed to test whether the child can entertain the point of view or beliefs of another when they contradict her own true beliefs. There are many variations on this experiment, the common thread to all is that the child, along with a confederate, both watch as a puppet hides an item in one location. The confederate then leaves the room and the child watches as the puppet removes the item from its original location and hides it in a new one. The child is then asked where the confederate will look for the item upon his return. There is a striking difference between the answers of the three year olds and those of the five year olds. Since the three year olds lack the ability to entertain a point of view other than their own, they mistakenly think that the confederate believes what they believe, that the item is now in a new location. The five year old, having acquired the ability to form representations, accurately reports that the confederate will look for the item in its original spot.

Gopnik claims that the changes that occur from the three year old's theory to the five year old's theory are an example of the same type of theory change that scientific theories undergo. Putting it in terms of the three steps of theory change outlined above, the child should start out with a non-representational desire-perception theory that provides adequate predictions for the child's own point of view, but not for that of others.

After making a series of inaccurate predictions about the desires and perceptions of others, and explaining these mistakes with ad-hoc hypotheses, the child should go through a transitional period from the old theory to the new representational theory, where she sometimes makes accurate predictions. The transition to the new theory is complete when the child makes very few inaccurate predictions. Concerning belief and the false belief task, the child is said to have a complete representational theory at the age of five, when she is able to entertain another person's point of view and consequently never fails the false belief task.

The general line of argument against Gopnik's account of changes in the child's theory of mind, led by Stich and Nichols, is that while her theory of the 'child as scientist' fits the data, other explanations or theories fit the data equally well (Stich and Nichols, forthcoming:32). They have thus thrown into question her overall strategy, which was to see how well the descriptions of theory and theory change, after being translated into cognitive science terms, would fit the data. Following from this overall objection, they cite major problems with her characterization of scientific theory and theory change, and this then throws doubt on her whole conception that the child is a scientist (Ibid:33-6). In other words, if scientists do not do science in the way she has construed it because her characterization of scientific theories and the way they change is incomplete, outdated or misconstrued, then there is little weight left to the statement 'children are scientists'. Of course, this objection would be rendered quite weak if Gopnik where able to show that her theory of the child as scientist is still plausible on other theories about how scientists do science. She has nonetheless, failed to do this, and so Stich and Nichol's initial objection still stands, weak as it might be.

The way Stich and Nichols view the situation, Gopnik has got a much too narrow and outdated view of how scientists go about the activity of science. They further take issue with her conception of theory change, arguing again that it is narrow and dated, and that there are a good many other theories as to how theory change comes about that she has failed to incorporate or acknowledge. Aside from the Popperian view that she uses, there are other views, those of Kuhn and Feyerabend, for instance, that claim that subjectivity, society, or whim has a much larger role to play than Gopnik allows for. She has also, according to them, left out a whole contemporary period in the philosophy of science where relativism is claimed to play a large role in when theories will change and what new alternative will become the dominant explanation. It is also unclear how the child may be construed as a scientist on these other theories of science.

Perhaps Gopnik would have been better off if, to put it bluntly, she had never stepped into the territory of philosophy of science. By claiming that children are little scientists, or more recently, that scientists are big children, she has unwittingly placed the focus of her theory, which otherwise has quite a bit of explanatory power, on her underlying conception of theory as scientific in some sense. Once she has deemed her theory as akin to a scientific theory in some sense, she has stepped into a realm where it is still not agreed upon just how it is that scientists go about the business of science, as well as what, exactly, constitutes a scientific theory.

The story thus far is that Gopnik has made a worthy attempt at offering a version of theory-theory that satisfies the constraint mentioned earlier, that in order for theory-theory to claim simulation theory as one of its versions, the latter must, at the very least incorporate a sense of the term 'theory' that is clear and narrowly defined, such as Einstein's theory of relativity, rather than something akin to 'a fanciful speculation as to the way things might work'. However, because her theory incorporates such a narrow construal of the term 'theory', she runs into problems stemming from that very construal, namely, that it is the classical construal of theory, which is generally thought to be an incomplete sense of the term.

2. Stich and Nichols: Defenders of the Faith

As indicated by the title of this section, Stich and Nichols defend a general version of theory-theory rather than a particular version that incorporates a definite construal of the term 'theory'. It is interesting to note that they can criticize a version that does use a concise definition of the term, such as Gopnik's, without considering that the same type of objection could be made toward them. Here, however, the issue is the opposite: their construal is so vague as to be meaningless. The point of this chapter is to show that at the very least, there is no reason to believe that simulation theory is just another version of theory-theory, since the version that Stich and Nichols broadly defend does not even satisfy the constraint on the construal of the term 'theory' mentioned earlier. Add this to the fact that aside from defending a version of theory-theory that is broad in its construal of 'theory', and perhaps tacit in nature, Stich and Nichols take a stand on very few other facets of theory-theory that are defensible. Taking these points together, it would seem that there is less and less reason to endorse the version of theory-theory that Stich and Nichols defend, among other reasons because it is not an identifiable version, but rather a group of versions taken together. This tendency of theirs to endorse a general form of theory-theory, instead of a specific version that is clearly formulated, would be better suited to a discussion where simulation and theory-theory are considered "...the only two games in town." (Stich and Nichols,1995:90) A persuasive argument to the effect that this is not the case is the fact that there is much in-fighting going on within each side, and also that it has not been adequately demonstrated both that simulation and theory-theory exhaust the field, and moreover that the final explanation of our folk psychological capacities could even be a mix of the two sides.

As mentioned above, Stich and Nichols object quite strenuously to the narrowly construed sense of theory and theory-change that Gopnik has advanced. It is no surprise that they place themselves at the opposite end of the continuum, where 'theory' is construed most broadly and inclusively. Here is how Stich quotes himself as once having put it (in his1983 book) as well as how he and Nichols presently construe 'theory':

> Others, including one of the authors of this paper, have taken an even more permissive view. Our use of commonsense psychological terms, Stich once claimed, "is governed by a loose knit network of principles, platitudes and paradigms which constitute a sort of folk theory." (Stich,1983,p.1)
> And in more recent work, where the focus was the plausibility of off-line simulation theories, we have used the term 'theory-theory' in a way that would count just about any collection of beliefs as a theory. (Stich & Nichols,1997:4).

Stich and Nichols are thus explicitly acknowledging that their construal of theory-theory is so wide as to include a very loose sense of the term, i.e., Felipe Alou's theory about baseball, or my theory about rush hour transportation. It is then clear and obvious that the version of theory-theory they defend does not meet the constraint mentioned at the beginning of the chapter.

Stich and Nichols, thinking that they are in competition with the other side rather than with other versions of theory-theory, only take a stand on a minimum of issues regarding their version of theory-theory. One of these is that the theory is tacit, although they are again quick to point out that they do not take a stand on whether they think the theory is a sentence-like or rule-based system, or whether it is neither sentence-like nor rule-based (Stich & Nichols,1995:133). It would then appear that Stich and Nichols could at least take a stand on whether the theory is sentence-like or rule-based or not, and not worry about whether they are committed to defend a version of theory-theory that is subject to damaging objections. One is thus left to wonder at their recalcitrant attitude when it comes to taking a stand on how the theory is organized in the mind, and further to wonder why they aren't quick to vocalize their view, when a claim could be attributed to them that they do not endorse. For instance they could end up being labeled as subscribing to a view that they find repugnant, perhaps a language of thought, just to name an example.

More importantly than the unwillingness on Stich and Nichols' part to defend any identifiable claims made by their version is the idea that their version does not satisfy the 'theory' constraint mentioned at the beginning of this chapter. They cannot realistically claim that simulation theory is a version of their preferred view when they do not even put forth a minimally defined sense of 'theory'. On their account, any set of phrases, statements, ideas or platitudes constitutes a theory, and so on this view, anything from a newspaper article to lecture notes could be considered a theory. Moreover, on their very broad and inclusive notion, the argument could conceivably be made that all of our cognitive capacities depend or rest on a theory.

In order for Stich and Nichol's claim that simulation theory is another version of theory-theory to be taken seriously, they would have to tighten up their notion of 'theory' in such a way that it would then satisfy the 'theory' constraint. Gopnik's use of the term as it occurs in a particular era in the philosophy of science is not a good candidate, for they themselves argue that it is too narrow and outdated in the sense that it is no longer accepted as the definitive notion of theory in the philosophy of science. There is also an explicit unwillingness on their part to take a stand on how 'theory' is construed, other than in the most wide and inclusive sense possible. There is thus no sense in contemplating what the situation would be like if they were to adopt a more restrictive sense of the term, for instance, a contemporary sense of theory as it would occur in recent philosophy of science. I don't think it is even possible to come up with a notion of scientific theory that is agreed upon by everyone, especially in the face of so many different conflicting viewpoints within the field as to what constitutes a theory and how theories change. It would be interesting to see how their version might fare if they were to still claim that our folk psychological practices rest on a theory, yet spell out in detail what notion of scientific theory they mean by the term.

Cognitive penetrability is one of the main issues that Stich and Nichols have been deep in discussion with the simulation side about. Generally, it has to do with the predictive power of one's set of folk psychological platitudes. On the account that Stich and Nichols defend, as well as most other versions, the prediction is that the theory will make incorrect predictions about a person's behavior if the theory does not contain a particular twist or variation on a general rule of thumb. For instance, experiments have been performed that illustrate a strange quirk of human nature: the Langer effect. Experiments have been conducted where an unsuspecting person 'off the street' is asked to rate an array of items in terms of quality that, unbeknownst to them, contains items of identical quality. Even though all items are of identical quality, most of the people pick out an item on the right-hand side. This tendency to pick out the rightmost item is the Langer effect. When test-subjects are told about the protocol in an experimental situation and asked to predict what the people off the street will do, they always make a wrong prediction, i.e., they consistently predict that the people will randomly pick any item. The question is: how does simulation and theory-theory account for this effect, and which side offers the better explanation?

Stich and Nichols claim that while theory-theory can account for the effect, simulation theory cannot, and thus theory-theory gives the better explanation for the effect. The reason why the test subjects made the wrong prediction about the the people participating in the test is because their folk-psychological theory lacks this nuance or twist to the rule of thumb "people will generally pick out the higher quality item from an array that contains items of different quality." The difference in this experiment is that the items are of the same quality, leading subjects to display the quirk, " when the items are of identical quality and you aren't aware of this fact, pick the rightmost item." The point is, as Stich and Nichols argue, that the folk theory of most people does not contain this strange platitude, and so they wrongly predict what the person will do in the situation.

Stich and Nichols further argue that simulation theory cannot account for the Langer effect, but I believe that they have overlooked the merits of the simulation theory, for in addition to accounting for this effect, simulation theory offers a further virtue concerning this issue, that of context sensitivity, that the theory-theory does not contain.

One could make a preliminary objection concerning the experimental design of these studies, claiming, as Harris does, that the experiments were designed to block the use of the simulation strategy. One could claim, particularly concerning the false belief task where the children are asked where they think that another person will think that an item is hidden, that this task would be difficult to perform if children simulate rather than consult a theory, simply because this type of meta-simulation is the most complex, and develops last in the child (Harris,1992:209). The results of this experiment are different if the child is merely asked to physically show where the confederate will look for the item. One could also make the case that experiments demonstrating quirks such as the Langer effect are designed to get at the causes of people's behavior, i.e., the causal connection between stimuli and their behavioral effects and it could be argued that we cannot logically have access to these causes (Farthing,1992:159).[2] Moreover, one could use Gordon's line of argument: as mentioned in the introduction to Gordon's version, he believes that simulation imitates real life. One way to distinguish simulation theory from theory-theory is that simulation theory is a practical or live or active process whereas theory-theory is a theoretical activity where the interpreter merely picks a platitude or a theoretical statement from his theory and applies it. Because simulation imitates life, one could argue that simulation is thus designed to pick up on the nuances of life, not like theory-theory, which goes only according to the theory, and is wrong if the theory is wrong or incomplete.

One could also argue that simulation theory is indeed cognitively penetrable, that is, that the simulation process sometimes generates wrong or inaccurate predictions. The general argument in defense of simulation theory on this point is that if wrong pretend inputs are fed in, the resultant output will also be wrong or inaccurate, thus the simulation process is indeed cognitively penetrable (Gordon,1992:176-7). But there is a better way to argue that simulation theory is cognitively penetrable, and this could be considered as a virtue that theory-theory does not have. One should rather argue that simulation theory is context sensitive, that changes to the target's situation, such as objects to be avoided or things to be afraid of that suddenly appear, which influence or change the target's behavior will also be picked up by the simulator. On the theory-theory account, subtle changes in the situation will not affect predictions because the predictions come from a static theory of platitudes.

This is exactly the point of experiments demonstrating the Langer effect. These experiments are designed to illustrate departures from the general rule of thumb, as mentioned above. In cases like this, theory-theory will generate wrong predictions because the theory is lacking a particular crucial bit of information, or a departure from a general rule of thumb. It will also generate wrong predictions, however, if the situation that the target is in changes slightly during the episode, whereas the simulation process will pick up on these changes and still produce accurate predictions or interpretations, because simulation imitates real life. Gordon, the author of the phrase 'simulation imitates real life', has developed precisely the type of process, partial projection, that is designed to pick up on subtle changes in context that sometimes alter the outcome of simulation of another's situation.

I began this chapter with the purpose of convincing the reader that theory-theory is a tired alternative that has run out of steam and should now hand over the flag to someone else, namely simulation theory. This thread will continue into the third chapter, where I consider each side's stance on the question of self-knowledge. I hope that the reader has been convinced at least to place less faith on the theory-theory's robustness, especially considering how poor a general defense of theory-theory turns out to be, i.e., on such a wide construal of 'theory', the claim that simulation is just another version of theory-theory turns out to be an empty claim. On the other side of the issue, i.e., a well defined version of theory-theory, things are not much better in the sense that the author must provide a very specific type of proof for the particular construal of theory chosen. In this chapter I have committed a bit of philosophical ju-jitsu in that I have allowed Stich and Nichols to provide the means to shake the foundations of a member of their team, Alison Gopnik, instead of trying to dream up all the criticisms myself. The fact remains, nonetheless, that no matter how 'theory' is construed, narrowly or inclusively, simulation need not be considered as just another of its versions, and the theory-theory still does not amount to a theory with very robust claims.

ENDNOTES

[1] My purpose here is not to characterize the so-called 'classical view' of scientific theory as it has actually been characterized in the philosophy of science, but rather to describe the term as Gopnik has borrowed it in order to explain how children acquire the set of folk psychological capacities. I will thus be using the phrase 'classically construed' as a shorthand for her particular chosen construal of it, and not the construal as it occurs in the philosophy of science.

[2] This point is related to the issue of introspective access. Access to the causes of one's behavior is often conflated with access to the contents of one's mental states, both falling under the heading of introspective access. The claim behind introspective access is that one has direct experience only of the contents of one's current mental states. More will be said about this issue in the next chapter.

CHAPTER THREE

INTROSPECTIVE ACCESS: SIMULATION THEORY VERSUS THEORY-THEORY

In the first chapter I argued that simulation theory could stand on its own and should be considered as an alternative viewpoint in the debate about the nature of our folk psychological practices. In chapter two I supplied the most important reason for thinking that simulation theory is not just another version of theory-theory: the lack of an agreed-upon construal of the word 'theory'. Now that I have shown that simulation theory can stand on its own in some sense, I wish to submit the two rival theories to a test. The issue here is fundamental to any explanation of a theory of mind, that is: how do we go about ascribing mental states to ourselves as well as to others. The test is related to the three fundamental differences between Gordon and Goldman/Harris, discussed in chapter one. As we saw, Gordon's version of simulation does not make the claim of analogical inference, since this in turn requires that one be able to reliably identify one's own mental states, and Gordon does not wish to include any type of introspective access claim in his account. As we shall see, Gordon does indeed account for the capacity to ascribe mental states to oneself and to others, but in a quite different manner than Goldman and Harris. The general point of this chapter is to determine how well each side does at explaining the fundamental capacity that ordinary folks have of ascribing mental states to themselves and each other. My aim is to convince the reader that simulation theory (particularly the account of Goldman/Harris), perhaps partly because ascription of mental states to ourselves and to others is integral to its theory, offers the better explanation for how we identify our mental states and those of other individuals.

Due to the nature of the topic of self-ascription and its link to introspective access and self knowledge, some terminological clarification is required as well as the establishment of a general constraint, before discussion can begin. The constraint has to do with the type of account of our folk psychological practices that is of concern in the present discussion. There are two types that, if conflated, could result in the situation of two authors talking past one another for each type seeks to answer a question at a different level of explanation. The first type is the 'folk' psychology account, that perhaps authors such as Paul and Patricia Churchland refer to when they speak of the account being mistaken and eventually replaced once we know more about the human mind and its apparatus. The other type of account, more scientific or neuro-biological in nature, is perhaps what the Churchlands hope will eventually replace the mistaken 'folk' account. But it is the 'folk' account that is relevant to the present discussion, because it attempts to explain folk psychological practices as they occur at the pedestrian commonsense level, whether such practices are mistaken or not. I would thus like to impose the constraint on the account of our folk psychological practices such that it explains our commonsense intuitions about what we do when we ascribe mental states to ourselves and to others. Goldman also imposes a similar constraint in his discussion, although he uses a different terminology: he thinks that the account of folk psychological practices should be "psychologically realistic" (Goldman,1993:16).

Some clarification of terminology is also required, related to the concepts of 'self-ascription', 'introspective access', and 'self-knowledge'. I take it that self knowledge has a wider meaning than both introspective access and self-ascription. Self knowledge as I understand it here includes, in addition to the identification of mental states in oneself, attributions of personality traits and other insights not necessarily relevant to the identification of mental states in oneself. I take introspective access to be the act of directly identifying one's mental states, in the sense of 'looking inward' and will not be using this term except in the discussion on Goldman's view. The reason for doing so in the context of the present discussion is as follows. Although it might seem plausible to assume some type of introspective access in the identification of mental states in oneself, to do so would be to beg the question in both Gordon and Gopnik's account, for they deny that self ascription of mental states relies on any type of introspective access. I prefer to use the more neutral terms 'self ascription' and 'other ascription' to mean the identification or labeling of mental states in oneself and in others, respectively.

I want to mention here that I will not be assuming any type of taxonomy of mental states, for that would again be begging the question. Rather, it is the job of each side to prove either that mental states are organized into a theory according to their causal roles, according to the theory-theory side, or that mental states are distinguished by their quale or defined by the conscious experience of the person, according to the simulation side. (Goldman,1993:24)

My strategy in this chapter is to examine the accounts of Gopnik on the theory-theory side and those of Goldman/Harris and Gordon on the simulation side concerning the issue of self-attribution of mental states. The goal is to determine which side gives the better account of how ordinary folk go about identifying their own mental states as well as those of other individuals.

One could make the preliminary objection that the simulation side has a clear advantage in this issue since self ascription, in fact introspective access in the case of Goldman's view, is an integral part of the theory. This version of simulation should thus already have a reasonably robust explanation for how the process works and one would be initially inclined toward that explanation. One could further object that the situation is set up in favor of the simulation theory in the sense that there is a constraint on the sense of 'folk psychological practices' understood in this discussion that further gives simulation theory an edge over the theory-theory. However, since the issue in this chapter is about evaluating the accounts of Gopnik, Gordon and Goldman/Harris on their explanation of the practice of self-ascription as it is integral to our set of folk psychological practices, I don't think that this is begging the question.

The form of discussion that has typically gone on in the current debate between theory-theory and simulation theory is the following: Each side makes a different prediction concerning the issue at hand. For instance, one possible question where the two sides make different claims is the developmental question of how the theory of mind evolves in the child. The theory-theory side explains it as analogous to a change in a scientific theory whereas simulation theory (Harris in particular) explains it as an increase in imaginative flexibility in the child (Goldman,1993:27). Concerning the present issue of mental state ascription, each side makes a quite specific and different developmental prediction concerning the order of development of the ability to ascribe mental states. Theory-theory predicts that the capacity to make self and other attributions will arise at the same time, since both types of attributions are on a par: in both cases a theory is being deployed. The simulation side, or the Goldman/Harris version, predicts that self-attributions must precede other-attributions since attributions of mental states to others are based upon an analogy made on the basis of identifications of mental states made in oneself. There is thus an asymmetry in the simulation prediction whereas in the theory-theory, mental state ascription to oneself and towards others arises simultaneously almost as a single ability. The way the debate unfolds is that these different predictions are checked against the empirical data, with a view toward determining which side's prediction better fits this data.

The situation is obviously not as cut and dried as it may appear above. Aside from constraints that need to be made on terminology (mentioned above), there is also the consideration that there are slight variations from author to author on each side's prediction concerning mental state ascription. For instance, on the theory-theory side, in addition to arguing a slightly different version of the general theory-theory claim that mental state attribution and identification is based on a theory, Gopnik also argues that the claim that we are in direct unmediated contact with our mental states as adults, in other words, introspective access, is an illusion. Some theory-theorists, notably Carruthers, do make the case that we have introspective access to some of our mental states, namely the occurrent ones (Carruthers, 1996:35) On the simulation side, Gordon's claims concerning the issue are also at odds with most other simulation theorists, since he is the only author who does not make explicit use of the notion of interpreting another person's behavior based on an analogy from the identification of one's own mental states, and he denies that self ascription involves introspective access. Following from these differences, it is useful to view Gopnik's account as falling slightly off the beaten path from the general theory-theory account, and also Gordon's as different from the simulation account. It is this contrast within accounts on each side of the debate that helps to make the discussion on mental state ascription sharper and more representative of the overall debate.

Since there are minor differences in each account even within one side of the debate, for instance between Gopnik's and Carruther's predictions on the theory-theory side, one would assume that favoring one account over another should be easy and that the data should be able to support one account or another. However, one problem with 'letting the data decide' is that often the data does not appear to fit or argue for either rival hypothesis that is concerned. In Gopnik's case, according to Stich and Nichols, the problem was precisely that her theory as well as numerous other theories fit the data equally well. Yet a further problem that is related to this is the concern first raised by Harris that some experiments are designed with the purpose of blocking the use of one or the other side's strategy, or similarly that the experiment contains erroneous or just plain bad methodology (Harris,1992:35). For these reasons, I would argue that one should be cautious in interpreting experimental results, as well as making generalizations based on one or two sets of experimental data.

Keeping the constraint mentioned above in mind, as well as the cautionary remarks about experimental data, let us now consider first the theory-theory explanation for self ascription, in particular the account of Gopnik, and in the following section I will consider those of Goldman/Harris and Gordon from the simulation side.

1. Self Ascription and the Theory-Theory

As mentioned, the general claim from the theory-theory side regarding the issue of mental state ascription is that mental states are attributed to oneself or attributed to others on the basis of a theory of mental states that is used whenever the need arises, and possibly only tacitly so. To attribute a mental state to another requires that a person understand and have mastered the concept of the mental state in question. This understanding in turn depends on understanding the overall theory in which the concept plays an explanatory role, as a result of the causal links that occur between this mental state and others, and between stimuli and mental state and mental state and behavior, within the folk psychological theory. Stich and Nichols, and Carruthers claim that a large part of the folk theory will have to have been already assimilated in order for the individual to be able to identify or attribute mental states. It is the lack of a particular pertinent fragment of the theory that causes the subject to make erroneous predictions about the mental state or behavior of another person. The child's theory of mind presumably changes as more and more fragments are assimilated, resulting in the fact that the child can, for instance, attribute a false belief to someone at the age of five, but not at the age of three. The developmental prediction concerning the order of appearance of mental state ascription is that the child will be able to identify her own mental states at the same time as she will be able to attribute mental states to others.

That being the general theory-theory account of how we are able to attribute mental states to ourselves and others, we are now in a position to ask the question: How at odds is Gopnik's version of events and at what point does her theory depart from this general picture? As noted in chapter two, Gopnik's account is the strongest version of the theory-theory in the sense of her construal of 'theory' as scientific, and this puts her into a category of her own much like Gordon. Concerning the general developmental prediction that the theory-theory side makes, Gopnik makes the same general claim, that the capacity to make self-ascriptions arises at the same time as the capacity to attribute mental states to others. She also agrees with the general theory-theory claim that children deploy a theory in ascribing mental states. Additionally, most other theory-theorists also agree with her claim that the adult's theory is vastly different from the child's, and that the child thus undergoes a quite marked theory change at some point.[1] Overall, Gopnik makes two general claims, only one of which, to my knowledge, other theory-theorists agree with. The first is this idea of theory-change in the child's theory of mind, which enjoys almost general consensus among other authors. The other is more controversial, and is a direct challenge to the commonsense intuition that we are in direct contact with our mental states.

According to Gopnik then, the child's first theory of mind, before it has changed into one of an adult, is in error about many things. Here is how she states it:

> As young children we have psychological states, we have psychological experiences, and we have beliefs about our psychological states. Our beliefs about at least some of these states, however, are consistently incorrect and differ from the beliefs we will have later... Young children do not

> seem to believe that their own psychological states are intentional, nor do they experience them as intentional, in the way adults do. Since we were all once such children, what we think we know about ourselves changes radically. (Gopnik,1993:2)

Here Gopnik is referring to the radical change in the theory of mind that occurs somewhere between the ages of three and five. If children have a theory of mind that is radically different from that of adults, then it follows from this, according to Gopnik, that some of the child's beliefs about herself could be said to be mistaken in some sense, if one retrospectively compares the adult theory to the child's.

Gopnik goes on to argue the second point that the adult's intuition that she experiences her mental states directly is mistaken in the same sense that some of the child's beliefs about the child's own mental states are. To put it in other words, she argues that there is a similarity between the child's beliefs about her theory of mind and the adult's beliefs about her theory of mind. Her claim is specifically the following: children don't believe that they have intentionality concerning mental states, whereas adults believe that they have direct experience of their mental states. Both of these intuitions or 'beliefs about beliefs' are mistaken, on Gopnik's account.

Gopnik explains the illusion that we adults have of being in direct contact with our mental states as analogous to the illusion of the expert. The novice-expert scenario is often referred to in educational settings to explain how the x-ray student, for instance, goes from being the awkward novice to the expert x-ray technician able to make immediate and accurate diagnoses (Azevedo et al,1997). Gopnik uses the novice-expert analogy to explain away as an illusion the intuitive notion that we have direct experience of our mental states. On her account, the novice is the child who does, at times, experience mental states directly, and the adult is the expert who experiences a theoretically reconstructed mental state or intentionality rather than the mental state itself. The reality of the mental state ascription situation is that we do not have direct experience of our mental states, we rather have the illusionary feeling of being in direct contact, which is brought about by our expertise with the theoretical construct of intentionality. Her argument is that intentionality is a theoretical construct, invented by the child at around the age of four, and designed to bridge the gap between mental states and our experience of them. Adults may believe that they have direct experience of their mental states i.e., have introspective access, however they do not. What mediates between mental states and our experience of them is this construct known as intentionality, defined by Gopnik as "... complex states that mediate between beliefs and desires and actions." (Gopnik,1993:5) The novice part of the analogy, in other words that single episode where there was no intentionality involved, what Gopnik calls the 'Wolfian' or 'Joycian' stream of experience, does occur, she admits, but most often what we are in contact with is the theoretically reconstructed mental state rather than the phenomenal feel of a genuine occurrent mental state (Gopnik,1993:12).

Having made the concession that we do, at times, directly experience our mental states, Gopnik can no longer make the 'illusion of the expert' argument stick, in effect, arguing the stronger claim that the intuition that we have direct contact with our own mental states is illusory. In other words, she cannot smuggle in the concession that at some point, perhaps our first experience with a never-before-experienced mental state, we do experience the state directly and without mediation. For then, the objection could be made: At what point does the mental state change from a 'Wolfian' newly-experienced mental state, to a theoretically reconstructed mental state?

As Gopnik states it in her conclusion, she wants to suggest an alternative order to the intuitive account of experiencing mental states:

> The commonsense picture proposes that we have intentional psychological states, then we have a psychological experience of the intentionality of those states, then we observe our own behavior that follows those states, and finally, we attribute the states to others with similar behavior. I suggest a different

> sequence: First we have psychological states, observe the behaviors and the experiences they lead to in ourselves and others, construct a theory about the causes of those behaviors and experiences that postulates intentionality, and then, in consequence, we have experiences of the intentionality of those states. (Gopnik,1993:12)

In effect, she proposes a few changes to the commonsense version of events. Intentionality is no longer the first thing experienced by the person, contrary to what she thinks is the commonsense view. Intentionality is rather constructed by the person as a result of determining the behavioral response caused by the mental state. It is then experienced by the person as a last step.

Her general point is that the commonsense intuition "... that our knowledge of intentionality, like knowledge of sensations, comes directly and reliably from our psychological experience." is perhaps incorrect (Gopnik,1993:2). She believes instead that our knowledge of own mental states comes from the same source as our knowledge of mental states of others. She ends up claiming that although direct experience of psychological states does play a role in the initial construction of the mental state theory, the theory is thereafter applied without the phenomenal experience.

For Gopnik's argument to be consistent, she must allow that each different mental state that is newly experienced by the person must be directly perceived. Otherwise, the novice part of the novice-expert analogy, where presumably the new mental state is experienced for the first time, is missing. Additionally, Gopnik must provide a process for distinguishing mental states experienced for the first time from those mental states that are intentional and theoretical. If she doesn't, we need not be convinced that her account is different from the commonsense one, i.e., we can continue to believe that all mental states are experienced directly and that there is no subsequent dropping off of the direct phenomenal experience of the mental state, to be replaced by an experience of the theoretically constructed mental state, or intentionality.

On a more specific level, the fact that there is no asymmetry in Gopnik's account, i.e., she equates first person mental state ascription with that of third person, leaves the impression that there is no privilege to the identification of one's own mental states by oneself. Note here that some of the other theory-theorists merely state that both capacities arise at the same time, but Gopnik also states that self and other ascription both come from the same source. She makes the daring claim that this intuitive feeling that we have of enjoying immediate privileged access to our own mental states is an illusion, the mental states are actually theoretical constructions, based on the first experience of the mental state, which was direct or 'Wolfian'. It could be argued that Gopnik's account leads to a dissociative view of the individual, in that there is a separation between that part of the person that experiences the 'Wolfian' mental state, and the part that experiences the state as a theoretical construct. In this case there would be a core self that directly experiences mental states in a 'Wolfian' manner, while the other self reconstructs our mental states and those of others on the basis of outward behavior and then constructs a suitable candidate mental state based on the causal consequences of that behavior.

An objection could be made to Gopnik's view as to how we come to have knowledge of our mental states and those of others, underlining an inconsistency in the account. As mentioned, she places no more importance or privilege on self ascription than she does on ascriptions to others. But then why do we not enjoy the same high degree of intuition about the mental states of others as we do our own? Moreover, why don't we achieve the same degree of accuracy and reliability when we ascribe mental states to others as we do when we ascribe them to ourselves? If, by the time we are adults we have become experts at using mental concepts, why do we not have an equivalent sense of expertise and 'privileged' access when it comes to ascribing mental states to others?

The expert-illusion analogy that Gopnik uses is probably better suited to illustrate the distinction between what used to be known as declarative and procedural knowledge, rather than the way she is using it to explain our commonsense intuitions about mental state ascription. In the 1970's, when cognitive psychology was just beginning to be recognized, one of the dominant explanations of the difference between 'knowing how' and 'knowing that' was this declarative-procedural knowledge distinction. Generally, it used to be thought that declarative knowledge was of the type that could be articulated, knowing that the capital of Canada is Ottawa for instance, and procedural was of the type that could be demonstrated, but did not lend itself well to articulation, such as riding a bicycle or executing a tennis serve. The interesting thing about these two types of knowledge is that most instances of procedural knowledge were once instances of declarative knowledge. In other words, the expert x-ray technician, who cannot explain how it comes about that he can 'see' the cancer (procedural knowledge) has forgotten that he at one time was a novice who had to go through the check-list procedure one item at a time and make a diagnosis based on the results (declarative knowledge). The illusion that Gopnik refers to is the illusion in the expert x-ray technician that makes him think that he does not go through the whole answer check procedure, but rather that he just directly sees the cancer. Explaining the x-ray technician scenario in terms of the novice-expert distinction works quite well, but I don't think that the distinction lends itself well to mental state ascription and the explaining away of our commonsense intuitions about mental state ascription.

Given the constraint mentioned earlier concerning the level of explanation at which the account of psychological practices is aimed at, it could be argued that Gopnik's account of the illusion of direct access is a level of explanation more suitable to a discussion on the nature of scientific psychology, rather than the level concerned with how the folk explain their own and each other's behavior. The reason is precisely because she is making a claim that something is an illusion, and contrary to our commonsense intuitions. At any rate her account does not seem very plausible as an explanation of our intuitions. She is also postulating an extra step in a process that has otherwise been thought of as direct, thus equating the act of attributing mental states to another person with the act of identifying our own personal mental states. If Gopnik's account was right, classical terminology designed to describe the act of identifying our own mental states would have to change: we would no longer identify mental states in ourselves, we would rather attribute them or infer mental states to ourselves, because the process is now on a par with attributing them to others. We can also no longer be said to have privileged or direct access to our mental states anymore either, even though they occur within our own bodies. On Gopnik's account, we might say that we make 'objective inference' to our mental states.

For the reasons outlined above, and most importantly because Gopnik's account does not do a very good job at explaining our commonsense intuitions about the way in which we ascribe mental states to ourselves, I am not particularly convinced by Gopnik's account, especially by her claim that our intuition that we have direct contact with our mental states is an illusion analogous to one that an expert x-ray technician would experience in making diagnoses.

2. Ascent Routine: Gordon's Method of Mental State Ascription

The accounts of Gopnik and Gordon, although they are on rival sides in the debate as to the nature of our folk psychological practices, nonetheless have one element in common and that is that neither account postulates that we have direct access to mental states, i.e., in the introspective sense of looking inward. While Gopnik claims that direct introspective access is an illusion, Gordon steers away from the claim because he considers the notion of introspectively based access and the implications stemming from it too problematic to include in the version of simulation he defends. Both authors are thus interested in tracing a new path as to how we have knowledge of our mental states that does not include the idea that we are in direct unmediated firsthand contact with them.

As mentioned in chapter one, Gordon defends a version of simulation theory whereby the process of behavior or mental state interpretation is not based on an analogical inference made from the simulator to the target. He chooses not to include the analogical inference because of the implications arising from accepting the account, i.e., it would require that the simulator be able to directly identify his or her own mental states. Gordon would then have to claim that the simulator uses some type of introspective access to identify his mental states. Self ascription, on this account, in turn requires concept mastery, and although Goldman and other simulationists accept concept mastery as necessary for introspective access without also accepting the further claim that this would require a theory, Gordon wants to avoid the whole controversy and thus he defends an account that steers clear of the assumption.

If an individual does not identify his own mental states by means of direct unmediated introspective access then how does he or she identify mental states, on Gordon's account? Self-identification of mental states is done by what Gordon calls 'ascent routine', this being a term he has coined himself. It is a process whereby questions about the status of a person's beliefs or desires are answered at a lower semantic level, thereby turning them into questions about the content of the belief or desire itself, and not about the status of the belief or desire, i.e., about the relation between the believer and his belief. For instance the question 'do I believe it will rain tomorrow?' is transformed into a question at a lower semantic level, i.e., to an object level question, and to answer it I would simply ask myself 'will it rain tomorrow?'. The propositional attitude that prefaces such utterances is simply dropped, and it is really the content of the utterance and not the status of my belief that I then ask myself about. Gordon's method of ascent routining is similar to how Evans describes belief monitoring in the individual:

> I get myself in position to answer the question whether
> I believe that p by putting into operation whatever procedure
> I have for answering the question whether p. (Evans,1982:225)

By proposing the ascent routine, Gordon is able to completely bypass the need to deal with the issue of mastery of mental concepts, for it is presumably not necessary to consult nor have mastered one's concept of belief per se in order to answer the question of whether or not it is going to rain tomorrow. One could argue that prefacing such questions with 'do I believe' or 'do I feel' is just a superfluous linguistic convention and has nothing at all to do with the content of the question nor with the relation between the individual and his belief. However, is this true in all cases of belief monitoring? In other words, aside from linguistic formality, is there no real function at all for the relation that the individual has vis-a-vis his belief?

Further, is there no link between the content of the belief and the fact that it is a belief? If someone were to ask me whether I believed that Clinton would be re-elected or not, would asking myself the object level question "Will Clinton be re-elected" be enough? Or is the question asking for something more than just a 'yes' or 'no' answer? By this I mean to get at the types of questions where an emotional reaction gets triggered by a question about the status of one's belief, or an issue where one's opinions could turn out to be contrary to the facts. In both of these cases, asking oneself an object level question is not enough, one is really being asked about the status of one's beliefs in addition to the information asked in the object level question. For instance, a question about an emotional issue such as "Do you believe that citizens should have the right to bear arms?", is asking more than object-level information from the individual. The issue here is whether that extra knowledge, i.e., an emotional reaction, is constitutive of a mental state that might be triggered by such a question or whether it is constitutive of the wider realm that is self knowledge, in which case it would not be a relevant objection to Gordon's views.

Concerning the issue of attributing mental states to others, it is the context or situation that the target finds himself in that will provide one key in determining what mental state the target might be experiencing. The other key is an extra step that is required by the simulator, that he re-center his egocentric map onto the target, so that the simulator 'becomes' the target in the target's situation. In Gordon's words, this is a matter of "... simulating O in O's situation", as opposed to the Goldman/Harris account which would instead be to simulate oneself in O's situation (Gordon,1995:61). Once this step is complete, the act of attributing a belief or a desire to the target becomes redefined on Gordon's account: "... to ascribe to O a belief that p is to assert that p within the context of a simulation of O" (Gordon,1995:61). In this way, Gordon is attempting to account for the fact that a pain or belief belongs to someone and that the simulator must have integrated this fact in order to be able to ascribe mental states to another person. He is thus able to get around the requirement that the simulator have mastered the concept of belief in order to be able to attribute a belief to someone else.

What then might be the prediction, on Gordon's account, concerning the developmental issue of which arises first, (if any), the ability to identify mental states in oneself or the ability to identify them in others? Here Gordon makes a further distinction between what he calls uncomprehending belief ascriptions and comprehending belief ascriptions, where uncomprehending ascriptions are made without the understanding that the beliefs may be false (Gordon,1995:62). On this distinction, Gordon makes the claim that uncomprehending belief self ascription or identification will arise first in the child, and even before the capacity for introspective access (Gordon,1995a:62). The reason is that the child learns how to employ the linguistic form of belief ascription, that is, to preface all object level ascriptions that they utter with the phrase 'I believe' long before they realize that they are stating a claim about their belief repertoire, and long before they are able to identify their mental states.

So far, Gordon's prediction accords with that of Harris/Goldman, since they too believe that the capacity to identify mental states will precede that of ascribing them to others. As for the ability to comprehendingly identify past beliefs and desires in oneself, however, Gordon departs from the view of Harris/Goldman and claims that this ability appears last of all. Gordon thus places more emphasis on the division between comprehending and uncomprehending self ascription than on the distinction between self ascription and the attribution of mental states to others. Not wanting to rely on the problematic issue of introspective access understood in the standard looking inward sense, he instead uses his idea of ascent routines, and ends up with the interesting claim that comprehending self ascription of past true beliefs occurs last of all. For Gordon, the self-identification of beliefs coupled with the understanding of whether they are true or false requires that one take "... a vantage point outside one's present perspective, whether by simulation or by memory demotion, and then reflecting back from that vantage point on what one counts as fact pure and simple and relegating it to a perspectival status." (Gordon,1995:62-3). To sum up, Gordon's account is different from the Harris/Goldman account in that he makes a distinction within the category of self ascription. He distinguishes between uncomprehending ascription which is done by ascent routine and arises even before introspective access, as understood not in the inward looking sense, and comprehending ascription, which is not direct in the same way that uncomprehending ascription is.

It would seem, on an intuitive level, that the above description is a quite accurate reflection of what we would do, as ordinary folk, when checking the status of our beliefs about an issue or a set of facts and the like. It seems plausible to say that we take an objective or third-person perspective outside ourselves and have a sort of conversation with ourselves, where we ask "'what is my position on this issue', or 'what do I believe is the case here.'" However, is it really necessary to posit this long third-person path where we determine our mental states through inference from overt behavior (Gopnik's view) or by taking a vantage point outside our present perspective (Gordon's view) when all along there is this short-cut (introspective access) that has much more intuitive appeal? As mentioned before, I think it is pretty much accepted that we cannot have direct access to the mental states of other individuals, mostly because we cannot get inside their heads, and so any ascription of mental states to others must be done by some type of inference. However, the same is not true for ourselves. For as a result of occupying our own bodies and thus heads, we are in direct contact with our minds/brains, and so why should we not have direct access to our mental states in the sense of looking inward? Gordon would perhaps respond to this by claiming that we do have direct access to our mental states, but only of the primitive uncomprehending type, gotten by ascent routines. Even if we grant that ascent routes provide direct access to mental states in the self that are of an uncomprehending type, why would Gordon then posit that comprehending self ascriptions are gotten from a third person perspective? Both types are relevant to self-ascription in the same person, and so one is left to wonder why they take different routes, since the only reason behind the distinction is the fact that one type is more sophisticated than the other.

I think that there is some plausibility to Gordon's account, in particular the claim he makes according to which children are trained to use the linguistic form of belief representation, i.e., to preface belief ascriptions with the phrase "I believe that ", even though they probably don't have even a moderate understanding of what it is to espouse a belief, much less a false one. However, I still wonder, as I did in the first chapter, why Gordon must end up splitting the process of self-ascription into two, thus ending up with the strange prediction that belief self ascription of the comprehending sort is the last to develop, instead of just going with introspective access.

3. Goldman and Harris

So far I have been discussing authors who have what are considered strong versions of their respective theory, and who are considered radical within their side of the debate. Their accounts cause a substantial amount of 'in-fighting', where authors within the same side of the debate disagree with each other. This is the case for Gopnik on the theory-theory side as well as for Gordon on the simulation side. Goldman's account of mental state ascription is rather an alternative to the theory-theorist's view and not so much to other simulationist accounts, in that he puts forth his claims in response to objections made by the theory-theorist. For instance, he considers his account an alternative to the theory-theory view in that the process of interpreting the behavior of another can be conducted without having any knowledge of laws or principles that govern behavior, as required by the theory-theory view. The taxonomy or cataloguing of mental states is not based on the causal connections between mental states and behaviors, as with the theory-theory.

Goldman meets the constraint mentioned at the beginning of the chapter, that the account of our folk psychological practices should gel well with our commonsense intuitions and be concerned with explanation at the pedestrian level. It can be stated with assurance that his account meets the constraint, since his own discussion also makes use of it. (Goldman,1993:2)

As mentioned in chapter one, Goldman uses the term 'empathy' in the broad sense of the term to denote the process of imaginative projection, which is just the process of simulation, that is, putting oneself into the shoes of another. However, it is his use of the term 'empathy' construed in the narrow sense that bears upon the present issue of mental state ascription. Generally, Goldman uses the term to describe the simulation or "... 'mimicking' of one person's affective state by that of another." (Goldman,1992:198). He describes the process as follows:

> Such initial pretend states [that are fed into the inferential or other cognitive mechanisms] are then operated on by psychological processes, which generate feelings, attitudes, or affects that are similar to, or homologous to, the target's individual states. Furthermore, just as the simulator is generally aware of his states as simulations of the target, so the empathizer is presumed to be aware of his vicarious affects and emotions as representatives of the affects and emotions of the target. (Goldman,1992:198)

In the case of the attribution of emotional states, on Goldman's account, once the pretend mental states are fed into the cognitive mechanism, they should spark off an emotion or sentiment that is similar in degree to that of the target. This seems to be a much more plausible account of how emotional states are identified than the theory-theory view, which describes it as a third person objective process that offers no way of determining isomorphism of mental or emotional states between the target and the interpreter. It is my opinion that this account is very plausible indeed, because it makes use of this aptitude that we all have of empathizing through 'remembering' a similar mental state what another person is feeling in a particular situation, because we have felt the same way ourselves given the same or a similar situation.

As mentioned in chapter one, Goldman (and presumably Harris) is the only author of those discussed that claims that we have introspective access to our mental states that is of a direct 'inward looking' type. The issue of whether we have such direct access is a contentious issue and most authors cite the same set of experiments by

Nisbett and Wilson (1977) as proof for the claim that direct experience of our mental states does not occur. However, as it turns out, Nisbett and Wilson only deny that we have access to our higher order cognitive processes, that presumably give us information about the causes of our behavior (Farthing, 1992:152). As Goldman notes, Nisbett and Wilson admit that we do have direct access to many of our current mental states, for instance "The individual knows a host of personal facts; he knows the focus of his attention at any given point in time; he knows what his current sensations are and has what almost all psychologists and philosophers would assert to be "knowledge" at least quantitatively superior to that of observers concerning his emotions, evaluations, and plans." (Nisbett and Wilson, 1977:225). Here Nisbett and Wilson concede that we have some sort of direct access to our current mental states, that is of a different or superior quality to that which other people have access to. There is thus no proof whatsoever that we do not have introspective access, at least to the contents of our current mental states, since the authors usually cited in favor of disproving direct access actually admit that we do have it. What they claim that we don't have access to are the causal connections between stimuli and their behavioral effects, which would be obtained through access to our higher order cognitive processes, and is logically impossible, according to Kenneth Bowers (Farthing,1992:159). The reason is because we presumably can't be both behaving or reacting to stimuli and observing ourselves reacting to stimuli, in a sort of third person mode, at the same time.

As far as I can determine, Goldman does not make any explicit claims regarding how his theory would explain the developmental chronology of mental state attribution. It is possible to infer or extrapolate that his claims would follow those of Harris, since their accounts are quite similar. The added aspect of empathy in Goldman's account should not make too much difference, and should perhaps reinforce the claim that mental state attribution to others must develop after self attribution, since the simulator would otherwise not have any recognitional capacity for emotional states upon which to base his or her interpretation of the target on.

Although Goldman doesn't offer up a prediction per se concerning how ascription will develop chronologically speaking, he does have a ready explanation for the experimental results in the false belief task. His explanation is that, on the simulation view, the child must develop the ability to attribute beliefs to others, and realize that these beliefs may or may not be similar to her own. He explains the difficulty that children have with attributing beliefs to others that are different from their own as the as-yet lack of an ability to entertain another's point of view as different from their own. The child mistakenly attributes to the other her own beliefs concerning where the item is hidden because she has not yet acquired the capacity where she realizes that the beliefs of others might differ from her own, and further that the beliefs of others might still be held even though contrary to fact.

Harris does make an explicit prediction regarding this developmental issue, since he has much to say about the difficulty in children of calling up past mental states. Recall that his account of simulation is based on the concept of imaginative flexibility, and that the child's theory of mind becomes more sophisticated due to incremental increases in imaginative flexibility. The most difficult simulation to perform, and thus the last ability to develop, is that of simulating someone in the past tense who has beliefs that are counter to a real world situation, and this is because the simulation is twice removed from default simulation, which would be to simulate oneself or someone else in the present, and where there are no counterfactual beliefs involved. Harris's main concern and thus objection about the experiments involved in the debate is that they only inquire either about the child's beliefs in the past tense, or about false beliefs (counterfactuals).

His version of simulation predicts that self ascription of mental states should arise first, because the child has the shortcut of direct access to his or her mental states. This same shortcut should also form the basis for attribution of mental states to others, since it is this very mechanism, as well as the act of putting oneself, in imagination, into the situation of another that allows the simulator to read off the mental states that arise as a result of being in the imagined situation.

Having entertained the views of Gopnik from the theory-theory side as well as Gordon and Harris/Goldman from the simulation side concerning the issue of how we identify mental states in ourselves and attribute them to other individuals, it is now time to determine which view has the most plausibility. To reiterate the strengths and shortcomings of each in a brief manner, Gopnik's view defies our commonsense notion that we somehow have a more direct relation to our own mental states than to those of other people. She claims no asymmetry in the order of acquisition of the capacity for self ascription and the capacity for other ascription. Her view transforms self 'identification' into the more objective self 'attribution' and puts it on equal footing with attribution to others, claiming that both capacities stem from the same source. Further, she sees the whole intuition of introspective access as an illusion, the situation being analogous to that of an expert. The expert thinks he or she physically sees what is really a product of extensive past theoretical experience of the object. The same can be said of self ascription, according to her, mental states are not directly experienced, our knowledge of them is rather the product of an extensive theoretical history. This account might work well to illustrate the distinction between declarative and procedural knowledge, but does not seem a likely candidate for how we identify and attribute mental states, in part because it flies in the face of the commonsense intuition that we have a more direct relation to things that go on inside of our own heads than to things that go on outside of our bodies.

Gordon's account suffers from many of the same problems that Gopnik's account suffers from, since he chooses not to take the shortcut of claiming that we have direct access to our own mental states. The fact that he must go to such lengths in explaining the identification and attribution of mental states because he has no recourse to the shortcut of introspective access means that his account suffers from the same problem as Gopnik's that is, the theory defies intuition.

The Goldman/Harris view, in my opinion, offers the best explanation of how we identify mental states in ourselves and attribute them to others. In the absence of proof that we don't enjoy introspective access to our current mental states, his account offers the most simple and elegant explanation concerning our ordinary capacity to ascribe mental states to ourselves and each other. Goldman's use of the construct of empathy proves to be particularly apt, especially when one considers the large amount of explanatory power it affords simulation theory in the area of identifying and ascribing emotions. In fact, I would be influenced toward Goldman's account even if all it had to offer that was novel was this empathetic explanation for emotion, since on other issues such as the interpretation of behavior when the context is important, just about any form of simulation will do.

ENDNOTES

[1] Fodor (1992) argues to the contrary, that rather there is a continuity between the child's and the adult's theory of mind, and that the child does not undergo a monumental theory-change at any point. He explains the apparent difference between the child's and the adult's theory of mind as the child having limited access to the range of cognitive resources that the adult has at her disposal.

CONCLUSION

In this thesis I have argued the general point that simulation theory should be considered a strong contender in 'The Interpretation Debate' which is presently on-going in the contemporary philosophical literature. I have tried to give the reader a bird's eye view into some of the current issues in that debate, also by describing some of the important authors' views, so that the reader has an idea of what simulation theory and theory-theory basically represent.

I began by revisiting a question that many simulationists had already assumed an affirmative answer to: should the simulation theory be considered as a viable alternative to the theory-theory? The major objection from the theory-theory side was that simulation theory should be considered as just another form of theory-theory. The reasons why, which I dealt with in chapter one, were the following. The first objection was that if simulation is to postulate some sort of analogical inference from the simulator to the target, the simulator would need to be able to identify his own mental states, upon which the analogy is then based. To do this the simulator would have to have mastered the theoretical constructs that the mental states are an instance of. Concept mastery requires a background theory, in order to house and provide a structure for the theoretical concepts such as 'belief' and 'desire'. The objection is thus that if simulation is based on any type of self-ascription of mental states, then simulation theory rests on a background theory and thus it is just another version of theory-theory. A way to resist this conclusion is to argue, as I have, that there is not an established link between concept mastery and a background theory, and thus simulation theory can postulate introspective access without necessarily committing themselves to also accept the claim that this requires a background theory. My second aim in chapter one was to show how different Gordon's account was from both Goldman's and Harris's, in order to show that he is immune to the objection made by the theory-theory, since he does not postulate an analogical inference and is therefore not subject to all of the 'theoretical' objections that arise from it.

In chapter two I examined the most important component upon which the theory-theory objection rests: the construal of theory-theory that three of the authors employ in their accounts. I imposed a constraint, according to which the theory-theorists must employ a reasonably restricted sense of the term 'theory', otherwise the claim that simulation theory is just another version of theory-theory could be considered as an empty claim. As it turns out, none of the accounts that I examined had a reasonably restrictive construal that was uncontroversial, and I suspect that it would be quite difficult to find one, resulting in the conclusion that we need not be persuaded that simulation theory is just another version of theory-theory.

A test was the focus of chapter three, in which I examined accounts from both sides, with a view toward determining which side had the more plausible explanation for an integral capacity within our folk psychological practices: the ascription of mental states to ourselves and to others. In general, the simulation side offers the better account, and I am particularly partial to the Goldman/Harris version, mostly because it offers a simple and direct route to self ascription that fits with our intuitions about it. Another strength of the Goldman/Harris account is that it offers a more plausible explanation of how we ascribe emotions to others, that is, through empathetic identification.

It should be obvious to the reader by now that I believe that simulation theory is more than a viable contender in The Interpretation Debate. I hope that I have at least cast doubt in the mind of the reader concerning the robustness of the theory-theory's claims. I also hope that I have succeeded in converting at least a few readers over to the simulation side, because it offers a better commonsense account of our folk psychological practices.

There is one aspect of the debate that I have not examined in the previous three chapters, and that is the status of hybrid theories. Such theories claim that our folk psychological capacities depend on a combination or mix of simulation and theory-theory, and are endorsed by Jane Heal (1995) and more recently by Stich and Nichols (forthcoming). The main reason why I have left out discussion of this type of hybrid theory is that it is beyond the scope of the present thesis. To adequately examine the possibility of a mix requires that one delineate what the role of each side is in the explanation of our folk psychological capacities, and this could be the topic of another complete thesis. However, such an endeavour could be considered a natural extension to what has been discussed here.

References

Azevedo, R., Lajoie, S., Deslaulniers, M., Fleiszer, D., and Bret, P. (1997). Radtutor: the theoretical and empirical basis for the design of a mammography interpretation tutor. *In Artificial Intelligence in Education.* B. de Boulay and R. Mizoguchi (eds). ISO Press.

Blackburn, S. (1992). Theory, observation and drama. *Mind and Language,* 7, 187-203.

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

Carruthers, P. (1996). Simulation and self-knowledge: a defense of the theory-theory. In Carruthers and Smith.

Davies, M. and Stone, T. (eds.), (1995). *Mental Simulation: Evaluations and Applications.* Oxford: Basil Blackwell.

Davies, M. and Stone, T. (eds.), (1995a). *Folk Psychology and the Theory of Mind Debate: Core Readings.* Oxford:Basil Blackwell.

Evans, G. (1982). *The Varieties of Reference.* Oxford: Oxford University Press.

Farthing, W.G. (1992). *The Psychology of Consciousness.* New Jersey: Prentice Hall.

Fodor, J. (1968). The appeal to tacit knowledge in psychological explanations. *Journal of Philosophy,* 65, 627-40.

_______ (1992). A theory of the child's theory of mind. *Cognition,* 44, 283-96.

Goldman, A. (1989). Interpretation psychologized. *Mind and Language,* 4, 161-85.

__________ (1992). In defense of the simulation theory. *Mind and Language,* 7,104-19.

Goldman, A. (1992a). Empathy, mind, and morals. *Proceedings and Addresses of the American Philosophical Association,* 66 (3).

___________ (1993). The psychology of folk psychology. *Behavioral and Brain Sciences,* 16, 15-28.

Gopnik, A, and Wellman, H. M. (1992) Why the child's theory of mind really is a theory. *Mind and Language,* 7, 145-71.

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

Gopnik, A. (1993). How we know our minds: the illusion of first person intentionality. *Behavioral and Brain Sciences,* 16, 1-14.

Gordon, R.M. (1986). Folk psychology as simulation. *Mind and Language,* 1, 158-71.

___________ (1992). The simulation theory: objections and misconceptions. *Mind and Language,* 7, 11-34.

___________ (1995). Simulation without introspection or inference from me to you. In Carruthers and Smith.

___________ (1996). "Radical" simulationism. In Stone and Davies, eds. (1996).

Harris, P. L. (1992). From simulation to folk psychology:the case for development. *Mind and Language,* 7, 120-144.

___________ (1995). Imagining and pretending. In Davies and Stone.

___________ (1996). Desires, beliefs, and language. In Carruthers and Smith.

Heal, J. (1986). Replication and functionalism. In Davies and Stone (eds). (1995a)

______ (1995). How to think about thinking. In Davies and Stone (eds.) (1995).

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy,* 50, 249-56.

Stich, S. and Nichols, S. (1995). Second thoughts on simulation. In T. Stone and M. Davies, (eds.) (1996).

___________________ Forthcoming. Theory-theory to the max.

___________________ Forthcoming. Cognitive penetrability, rationality, and restricted simulation.