RuCCS TR-45
November, 1998

A Cognitive Theory of Pretense

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

Stephen Stich
Rutgers Center for Cognitive Science
Rutgers University
stich@ruccs.rutgers.edu

Technical Report #45
Center for Cognitive Science
& Lab of Vision Research
Psych Bldg Addition, Busch Campus
Rutgers University - New Brunswick
152 Frelinghuysen Road
Piscataway, NJ 08854-8020

A Cognitive Theory of Pretense

Shaun Nichols
Department of Philosophy
College of Charleston
Charleston, SC 29424

and

Stephen Stich
Department of Philosophy and Center for Cognitive Science
Rutgers University
New Brunswick, NJ 08901

Abstract

Recent accounts of pretense have been underdescribed in a number of ways. In this paper, we present a much more explicit cognitive account of pretense. We begin by presenting a number of real examples of pretense in children and adults. These examples bring out several features of pretense that any adequate theory of pretense must accommodate, and we use these features to develop our theory of pretense. On our theory, pretense representations are contained in a separate workspace, a Possible World Box which is part of the basic architecture of the human mind. The representations in the Possible World Box can have the same content as beliefs. Indeed, we suggest that pretense representations are in the same representational "code" as beliefs and that the representations in the Possible World Box are processed by the same inference and UpDating mechanisms that operate over real beliefs. Our model also posits a Script Elaborator which is implicated in the embellishment that occurs in pretense. Finally, we claim that the behavior that is seen in pretend play is motivated not from a "pretend desire", but from a real desire to act in a way that fits the description being constructed in the Possible World Box. We maintain that this account can accommodate the central features of pretense exhibited in the examples of pretense, and we argue that the alternative accounts either can't accommodate or fail to address entirely some of the central features of pretense.

1. Introduction

Pretend play in young children is so familiar and so natural that it's easy to overlook how remarkable and puzzling it is. The wonder of pretend play hasn't been lost on major figures in developmental psychology (e.g., Piaget 1962, Vygotsky 1967). Until recently, however, the capacity for pretense received surprisingly little attention in the cognitivist tradition. Only in the last decade has there been a sustained attempt to map out the cognitive mechanisms underlying pretend play. During this period there has been an explosion of conceptual and empirical work on pretense (e.g., Leslie 1987, 1994; Harris & Kavanaugh 1993; Harris 1991, 1994, 1995; Perner, et al. 1994; Currie 1995a, 1997; Gordon & Barker 1994; Lillard 1993, 1994). Though controversy abounds in this area, it is widely agreed that a theory which succeeds in explaining the capacity for pretense will shed light on the mechanisms that underlie many other cognitive capacities as well.

Much of the recent discussion of pretend play can be traced to Alan Leslie's attempt to give a cognitive account of pretense (Leslie 1987). Leslie maintains that there are important connections between the child's capacity for pretend play and the child's capacity for understanding other minds or "mind reading". Since children pretend and seem to understand pretense in others fully two years before they succeed on standard versions of the false belief task,[1] Leslie maintains that pretend play constitutes a strikingly early example of mind reading skills. (Leslie 1978, 1994) The selective impairments of autistic children, who show deficits both in pretend play and in their ability to solve mind reading problems like the false belief task, provides one important piece of evidence for the connection between mind reading and pretend play. (Baron-Cohen 1995).

[1] In the classic false belief test, Wimmer and Perner (1983) had children watch a puppet show in which the puppet protagonist, Maxi, puts chocolate in a box and goes out to play. While Maxi is out, his puppet mother moves the chocolate to the cupboard. The children are asked where Maxi will look for the chocolate. Until about the age of 4, children fail this and similar tasks.

Although most researchers agree that there is an important connection between pretend play and mind reading, there are quite fundamental disagreements about the mental mechanisms that make mind reading possible, and pretense has played an important role in these disputes. One of the most heated debates is between advocates of the theory-theory and advocates of simulation theory.[2] Writers on both sides maintain that their theory provides a better explanation of mind reading phenomena and of the deficits seen in autism (e.g., Currie 1996, Carruthers 1996). As we'll see, simulationists have also argued that simulation theory provides the best account of the processes underlying pretend play (Gordon & Barker 1994). The stakes here are high, since if the simulation account of pretend play is right, it would mean that the basic mechanisms posited by simulation theorists have to be in place. And if that is true it might lend considerable plausibility to simulation theorists' claim that mind reading also depends on a process of simulation (Gordon & Barker 1994).

The capacity for pretense has also been awarded enormous importance in the recent literature for reasons that are quite distinct from the connection between pretense and mind reading. Simulation theorists have argued that the capacity for pretense is implicated in a wide variety of important cognitive capacities including counterfactual reasoning (Goldman 1992), conditional planning (Goldman 1992, Harris 1993), empathy (Goldman 1992, Currie 1995b, Sorensen 1997), and visual imagery (Currie 1995d). Simulation theorists have also argued that the capacity for pretense underlies imagination more generally (Currie 1995a,b). For instance, our imaginative encounters with fictions have been tied to pretend play or make-believe (Currie 1990, 1995a,b; Walton 1990, 1997), and Currie (1995c) has suggested that these same imaginative mechanisms are also used in moral understanding. Although much of this work has been pursued under the auspices of off-line-simulation theory (Gordon 1986; Goldman 1989, 1992), one needn't be sympathetic with off-line-simulation theory to think that the capacity for pretense is crucial for many of the above capacities. Indeed, though we have been frequent critics of off-line-simulation theory (Stich & Nichols 1992, 1995, 1997; Nichols

[2] For a good overview of this debate, see the essays in Davies and Stone (1995a & 1995b) and Carruthers and Smith (1996).

et al. 1995, Nichols et al. 1996), we think it's quite plausible that many of these capacities are intimately connected with the capacity for pretense. But before such proposals can be elaborated and defended, it's essential to have a plausible theory of the mental mechanisms that underlie the capacity for pretense itself.

Our goal in this paper is to offer such a theory and to compare it with other theories that have been proposed in the recent literature. The theory we will defend is a highly eclectic one which borrows ideas from writers on both sides of the debate between theory-theorists and simulation theorists. We don't think there is any principled way to classify our theory as falling into one camp or the other. Indeed, in another paper (Stich & Nichols 1997) we have argued that there *is* no principled division between these two "camps" and that it is high time that the labels "theory-theory" and "simulation theory" were retired. Though we won't argue the point further here, we think the theory of pretense we will defend buttresses the case for this conclusion by making it clear that there are plausible theories for which neither label is appropriate.

It is our contention that all the other theories of pretense that have been proposed in the recent literature are under-described in important ways, and in particular that all of them tell us far too little about the sort of mental architecture (or cognitive "boxology") that the theory is presupposing. As a result, as we'll argue in Section 4, it is often difficult or impossible to know exactly how these theories would explain one or another example of pretense, or how they would account for various aspects of the capacity to pretend. In an effort to avoid these problems, the theory we'll set out will be much more explicit about the mental architecture that the theory assumes, and about various other matters on which competing theories are silent. Though this approach has obvious benefits, there is a downside as well. For at a number of junctures there are several quite different paths that might be followed and, for the moment at least, there is little empirical evidence to indicate which is preferable. At some of these junctures we'll pause to consider a variety of options. More often we'll just make our best guess without stopping to explore the alternatives. Inevitably, some of these guesses will turn out to be wrong, and so too will our theory. But that doesn't really worry us, since it's our view

that the best way to make progress in this area is to develop detailed theories that can be refuted and then repaired as evidence accumulates, and not to rest content with sketchier theories which are harder to compare with the growing body of evidence. Being false, as the Logical Positivists often emphasized, is far from the worst defect that a theory can have.

Here's how we propose to proceed. In the Section that follows we will describe a number of examples of children's pretense drawn from various sources; we'll also recount some examples of adult pretense from our own studies. Then, in Section 3, we'll draw attention to some of the features of these episodes of pretense – features which, we maintain, a fully adequate theory of pretense must be able to explain. The list of features we assemble will thus serve as a sort of checklist against which competing theories can be compared. In Section 4, we will set out our theory of the cognitive mechanisms that underlie pretense, and show how the theory can account for the features on the checklist in Section 3. Finally, in Section 5, we'll sketch some of the other theories of pretense that have been offered and argue that our theory does a better job at explaining the facts.

2. Some Examples of Pretense

Before setting out our theory, it will be useful to offer several examples of the sort of episodes of pretending that we want our theory to explain. A familiar armchair strategy here is simply to make up some examples. But for a variety of reasons, we're not inclined to take that route. We worry that important features of pretense might be neglected if we limit ourselves to examples of our own invention. By looking at a number of genuine examples of pretense, we hope to get a much better idea of the richness and the range of the phenomena to be explained.

2.1. Pretense in children

Much of the literature on pretense is guided by two examples from the work of Alan Leslie. We will recount these examples below, because we too think that Leslie's examples are important to understanding pretense. In addition to the examples from Leslie, we will include two more examples of pretend play that haven't been discussed in the literature, but are genuine instances of spontaneous pretend play in children.

2.1.1. Pretending that a banana is a telephone

"How is it possible for a child to think about a banana as if it were a telephone...?" With this question Leslie (1987, p. 412) ignited the recent explosion of interest in pretense in cognitive psychology and simultaneously provided what is perhaps the most widely discussed illustration of pretense in the literature. There are really two different banana-telephone scenarios that often get discussed together. In one scenario, an individual child pretends that the banana is a telephone. For instance, a child might pick up a banana, hold it up to his ear and mouth and say, "Hi. How are you? [Brief pause.] I'm fine. OK. Bye." This kind of pretense will no doubt be familiar to parents of young children. Children also engage in pretense with others. For instance, a child and his mother might pretend together that a banana is a telephone. In an episode of this sort the mother may pass the banana to the child, saying, "It's for you." From a young age, children will play along in these pretend games, taking the banana and saying "Hello" into it. These scenarios get treated together by Leslie since he thinks that any child who can pretend that the banana is a telephone can also understand pretense in others. For our purposes, we want to maintain a clear distinction between individual pretense and group pretense.[3]

[3] Leslie has never used the banana / telephone scenario in his experimental work though he observed a number of spontaneous examples of the scenario in his own children. (Personal communication.)

2.1.2. Leslie's tea party

Another widely cited example of pretense also comes from Leslie's work. In a set of experiments, Leslie had children participate in a pretend tea party. Leslie describes the scenario as follows: "The child is encouraged to 'fill' two toy cups with 'juice' or 'tea' or whatever the child designated the pretend contents of the bottle to be. The experimenter then says, 'Watch this!', picks up one of the cups, turns it upside down, shakes it for a second, then replaces it alongside the other cup. The child is then asked to point at the 'full cup' and at the 'empty cup.' (Both cups are, of course, really empty throughout.)" (Leslie 1994, 223). When asked to point at the 'empty cup', two-year olds pointed to the cup that had been turned upside (Leslie 1994).

2.1.3. Monsters and parents

Our next example of pretend play in children comes from the CHILDES (Children's Language Data Exchange System) database (MacWhinney & Snow 1990). Though this database was initially established to study children's language it is an invaluable resource for studying a number of features of child psychology (see, e.g., Bartsch & Wellman 1995). In the following protocol, Val is 4 years and 7 months, and Abe is 4 years and 9 months (Garvey1979).

VAL: Pretend we're on the bus ok?

(VAL gets on, then off, gets on climbs on back.)

VAL: Enough room for two of us right?

ABE: Right.

VAL: You say hold on to me.

(ABE drives seriously while VAL talks. VAL puts arms around ABE to keep from falling off.)

VAL: Pretend we passed one and pretend there's a monster coming ok?...

ABE: No let's don't pretend that.

ABE: Cause it's too scary that's why.

VAL: Oh I don't think so....

VAL: Pretend ... we can play mother and father.

ABE: Ok and this was our car.

VAL: Ok. You're going to work.

(VAL gets off car as part of game; ABE follows her.)

VAL: Hey look look. We got this [a lunch box]; we got work tools.

(VAL speaks in high enthusiastic voice; VAL presents lunch box to ABE who seems to accept it as part of game; ABE takes tools to car VAL goes to stove, opens oven...)

VAL: We could cook this [shoes] for food.

(VAL holds up shoes to show ABE.)

(VAL puts shoes in oven, turns knobs, then takes them out and shows ABE.)

VAL: For food. Do you like chicken?

2.1.4. Dead cat

The final example of childhood pretense that we'll recount comes from a set of protocols of spontaneous pretense in children that was assembled by Rosalind Gould (1972). Gould reports that one three-year old boy on a jungle gym said, "I'm a pussycat. Meow, meow." He then came down from the jungle gym and lay on the ground, saying, "I'm dead. I'm a dead pussycat... I got shooted." (Gould 1972, p. 212).

2.2. Pretense in adults

There are a number of limitations to the available examples of pretending in children. However, since we want to develop a theory of pretense that can accommodate adult pretense as well as pretense in children, there's no need to restrict ourselves to examples of pretend episodes in children. Pretend episodes in adults will work just as well. As it happens, there are surprisingly few examples of adult pretense described in the

psychological literature. So we set out to collect our own. Our goal wasn't to design an experiment with a clever manipulation, and the kind of data we have doesn't admit of any sort of statistical analysis. Rather, we merely wanted to collect some genuine examples of pretense in adults.

10 college students participated in our little study. We asked individuals and groups of two to carry out a variety of scenarios. Each scenario was described on a separate sheet of paper. At the beginning of the session, participants were told to consult the written description and act out each scenario. For the individual pretend scenarios, the descriptions were as follows:

"Pretend that the banana (on the table) is a telephone."

"Pretend that you're home alone at night and you hear a suspicious noise in the basement."

"Pretend that you're a train."

"Pretend that you're a dead cat."

"Pretend that you're sleeping."

"Pretend that you're baking a cake."

For group pretend scenarios, the descriptions were as follows:

"Pretend you are in a fast food restaurant. Decide who will be the cashier and who will be the customer."

"Pretend you are in a fancy restaurant. Decide who will be the server and who will be the diner."

Eight of the subjects participated in the group pretend scenarios. Six of the subjects participated in the individual pretend scenarios. The order of the scenarios was always the same. Subjects were told that they could use whatever props were in the room. The scenarios were carried out in a lab room that had a telephone in it, along with typical office supplies. We supplied a banana for the banana/telephone scenario, and we provided the subjects with play money for the fast food restaurant scenario.

After subjects carried out the scenarios, we had an informal oral interview with them. We asked questions about what the subjects remembered from the scenarios, why

they behaved the way they did, and about the possible confusion between reality and pretense. For the group pretense participants, the interview was conducted with both participants at the same time. The four subjects who participated in both individual and group pretense were interviewed about their individual pretense with the other participant in the room. The remaining two individual pretense participants were interviewed alone. All sessions and interviews were videotaped, and the participants knew that the sessions were being videotaped. The videotapes were later reviewed to produce the transcripts.

In the sections to follow, we'll draw from the transcripts where appropriate. But before we conclude this section, we want to include two extended protocols, from which we will draw throughout the rest of the paper.

2.2.1. Fancy restaurant scenario, episode 1

The participants were two female college students.

B gets pencils and pens for silverware. Gets a piece of paper for napkin

A: Hi, how are you doing? (Hands B a syllabus as the 'menu'.)

B: Hi. Good how are you?

A: I'm fine. Can I get you something to drink to start?

B: Yeah, can I just have um, just water to start with. I'm waiting on someone. So.... He'll be a second.

A: Would you like to see a wine list or do you wanna wait until your...

B: No, that's fine.

A: You sure? OK. You want me to just come back in a few minutes?

B: (Looking at menu) Yeah. Thank you.

A: Ok I'll get that water for you.

A: (Gets cup from table, pretends to fill cup with water. Brings cup to B.) Still looking? Do you still want a few more minutes?

B: No, I think I'm ready to order. (Pretends to be reading from menu.) I'll have the chicken pasta without any pepper on it. Is there pepper in the sauce?

A: Yeah. I can do it without pepper that's fine. (Starts to write on pad, "chicken pasta".)

B: I'm allergic to pepper. I'll have the house salad with ranch dressing.

A: Oh, we don't have ranch. We have a vinaigrette.

B: Okay, yeah that's fine. And he should be here soon, but go ahead and bring him a house salad too. I'll wait till he gets here. I'll let him order. I don't want to order for him.

A: Do you still want me to bring the salad out?

B: Yeah That'd be fine.

A: Okay. I'll be back in a few minutes

2.2.2. Fancy restaurant scenario, episode 2

The participants were two male college students.

D: Pardon me I'd like to give you our specials for this evening. We have a grouper with a side portion of potato. As well as some cranberry juice poured on top as well as a great deal of greenery, I'm not sure what that is. Very pretty.

C: All right, that sounds wonderful. Let's see I had a question about the menu. The lobster tail is that pan seared or is it deep fried? Because I've got bad blood pressure I gotta watch it.

D: It is deep fried.

C: What about the lamb chops.

D: The lamb chops are exquisite.

C: Give me the lamb chops. Medium rare.

D: And for the woman.

C: I'll order for her. Give her the honey glazed chicken. And as a starter, let's see... (Pretends to pore over menu.)

D: I recommend the portabello mushrooms simmered in black oil.

C: I think I'm looking more at the herb crusted goat cheese. Oh, also, we were looking through the wine list. Um, which would you recommend out of these two Merlots?

D: The one from France.

C: The McGuigan brothers?

D: No, that would be the Scottish one. It's a Merlot from Champagne.

C: Oh, I've heard that's great. We'll have that one.

D: Certainly. (Pretends to leave.)

D: (Returns) And the food for the lady, for yourself . (Pretends to put food on table.)

There's your wine sir. (Pretends to pour into pretend glass.)

(C pretends to swirl the wine in glass, sniff, drink.)

D (Pretending to pour again.): Go ahead and have some more it's really good. Can I get you anything else?

C: Can you get me a sharper knife for the lamb chops.

D: Certainly. Anything else?

C: No.

D: All right. (Leaves)

D: (Returns with arms outstretched and pretends to be carrying a large object, pretends to hand large object to C.) Here's your knife sir. It's the biggest we have. I'm sure you'll find it's quite sharp. It is a two handed serrated blade sword of Japanese descent. Just be careful, we had a gentleman previously chopped...

C: (Pretending to look over the object, incredulous.) Did you take this off the wall?

D: It's the chef's. He's a Japanese man.

C: This is very ... pretty.

D: Well, it is specifically for chopping. You requested a sharp knife.

C: Well, thank you. (Pretends to lay down object on table.)

D: Can I get you anything else? Some bread?

C: Maybe a little freshly ground pepper?

D: Certainly. (Leaves)

D: (Returns, pretends to put peppercorns on table, pretends to chop them up with the knife.) Let me chop them up here.

C: Uh, ground. Do you have a grinder?

D: Well I really can't. I'll just, I'll just... (Puts foot up on table, pretends to crush the peppercorns with his heel.) Excuse me. I know that's a little crass. (Pretends to put the pepper on C's food.) Is that enough?

C: I think that's going to be about it.

D: Certainly. For the lady? Would she like some too?

C: Do you think I could talk to your manager?

D: (Pretends to be chopping more peppercorns.) Uh sure. (Turns, then whips around pretending to be waving the knife.) Oh, I've cut her head off. I'm sorry. That's a sharp knife. I'll definitely need the manager for this one. Oh, she's bleeding profusely. One moment please.

C laughs and indicates that he's finished with the pretense.

3. Some Features of Pretense

In this section we want to draw attention to a number of features that can be found in the examples of pretense we've recounted. Since they can be found in many other cases as well, they are features which a complete theory of pretense might reasonably be expected to explain. The theory that we'll offer in the following section won't have satisfying explanations for *all* the features on our list, though as we'll argue in Section 5, it can handle more of them than any of the other theories that have been proposed.

3.1. Getting pretense started: The initial premise

Typical episodes of pretense begin with an initial premise, an assumption about what is to be pretended. In Leslie's tea party experiments, the assumption is that the child and the experimenter are going to have a tea party. In our fancy restaurant scenario, the assumption is that one of the subjects is a diner in a fancy restaurant, and the other is the server. In the example from Gould, the assumption is that the boy is a pussycat. To get the pretense going the pretender must either produce the initial premise (if she initiates the pretense) or (if someone else initiates the pretense) she must figure out what

the initial premise is and decide whether or not she is going to proceed with the pretense. If the pretender decides that she will proceed, her cognitive system must start generating thoughts and actions that would be appropriate if the pretense premise were true. People don't always agree to accept the premise of a pretense, of course. In the example from CHILDES, Val proposes that they pretend there is a monster coming, but Abe refuses, because "it's too scary."

3.2. Inferential elaboration

Inference often plays a crucial role in filling out the details of what is happening in pretense. From the initial premise along with her own current perceptions, her background knowledge, her memory of what has already happened in the episode, and no doubt from various other sources as well, the pretender is able to draw inferences about what is going on in the pretense. In Leslie's tea party experiment, for example, the child is asked which cup is empty after the experimenter has pretended to fill up both cups and then turned one upside down. To answer correctly, the child must be able to infer that the cup which was turned upside down is empty, and that the other one isn't, although of course in reality both cups are empty and have been throughout the episode. In one episode of our fast food restaurant scenario, the subject who was pretending to be the cashier informed the "customer" that his order cost $4.85. The customer gave the cashier $20.00 (in play money), and the cashier gave him $15.15 change, saying "Out of $20; that's $15.15." In order to provide the correct change, the cashier must perform a simple mathematical inference. An adequate theory of pretense should provide an account of the cognitive processes that underlie these inferential elaborations.

3.3. Non-inferential elaboration (embellishment)

In addition to inferential elaboration, children and adults elaborate the pretend scenarios in ways that aren't inferential at all. In some instances, this is a matter of filling out the story provided by a scenario. For instance, in the protocol from CHILDES, Val, in playing the mother, pretends to cook chicken, using some shoes as props. However,

this elaboration is not logically or causally *entailed* by the pretend scenario. Similarly, in the first fancy restaurant episode we reported, the diner elects to have the chicken pasta for dinner. Some of these elaborations cohere with the decisions and choices that the person pretending would actually make. For instance, in the post-pretense interview, the diner in the first fancy restaurant episode said that she really would refuse a wine list because she doesn't like to run the risk of being asked for identification to prove that she's old enough to drink. In other cases, the elaborations depart from what the person would actually decide. The same subject claimed (in the pretense) that she was allergic to pepper, but in the post-pretense interview, she said that she just made this up. In another version of the fancy restaurant scenario, the diner behaved quite rudely, but in the post-pretense interview, he claimed that he would never behave that way in a real restaurant. More dramatically, in the second fancy restaurant episode we reported, the waiter pretended to decapitate one of the diners. This subject assured us in post-pretense interviews that in real life, he would never be so careless with a sword. A theory of pretense needs to be able to accommodate these kinds of elaborations as well as the more sober inferential elaborations.

3.4. Production of appropriate pretend behavior.

Perhaps the most obvious fact about pretense is that pretenders actually *do* things -- they engage in actions that are appropriate to the pretense. The child in Leslie's famous example takes the banana from his mother, holds it in the way one might hold a telephone, and talks into it. The adults who participated in our study did the same. In the fancy restaurant scenario, when the "diner" subject orders a meal some "waiters" wiggle one hand in a stylized imitation of writing, while holding the other hand as though it grasped a pad. Other waiters use a real pad and actually write down what the diner is ordering. The boy in the dead cat pretense that Gould observed lies on the ground, as a dead cat might, though his accompanying verbal behavior is not what one would expect from a dead cat, or from a live one. Adult subjects do much the same, though they are quieter. One adult in our study embellished the dead cat pretense by holding her arms up rigidly to imitate the rigidity of the cat's body after rigor mortis has set in. A theory of

pretense must explain how the pretenders determine what behavior to engage in during an episode of pretense. How do they know that they should walk around making jerky movements and saying "Chugga chugga, choo choo" when pretending to be a train, and lie still when pretending to be a dead cat, rather than vice versa? Equally important, an adequate theory of pretense must explain why the pretender does anything at all. What *motivation* does she have for engaging in these often odd behaviors?

3.5. Cognitive Quarantine: The Limited effects of pretense on the later cognitive state of the pretender.

Episodes of pretense can last varying lengths of time. When the episode is over, the pretender typically resumes her non-pretend activities, and the events that occurred in the context of the pretense have only a quite limited effect on the post-pretense cognitive state of the pretender. Perhaps the most obvious way in which the effects of the pretense are limited is that pretenders do not believe that pretended events, those which occurred only in the context of the pretense, really happened. A child who pretends to talk to Daddy on the banana / telephone does not end up believing that he really talked to Daddy. Moreover, as Leslie emphasizes, (1987), even very young children do not come to believe that bananas sometimes *really are* telephones. Nor, of course, do adults.[4] Moreover, even during the course of the pretense itself, what the pretender really believes is typically kept quite distinct from what she believes to be the case in the context of the pretense episode. Our adult subjects did not really believe that they were in a restaurant, or that they were baking a cake, or that they were dead cats. For example, we asked "At any time did it ever seem like you were really in a restaurant? Did you ever get confused and think that you were really in a restaurant?"; our subjects uniformly said 'no'. However, the pretender's belief system is not entirely isolated from the contents of the pretense. After an episode of pretense people typically have quite accurate beliefs about what went on in the pretense episode; they remember what they pretended to be the case.

---

[4] In our study of adult pretense we asked subjects who had pretended that a banana was a telephone (i) "Have you ever seen a real telephone that looked like a banana?" and (ii) "Have you ever seen a telephone that was made out of a real banana?" They all answered "No" to both questions.

Moreover, Gopnik & Slaughter (1991) reported that children find it much easier to recall what was pretended than to recall their own earlier false beliefs. Indeed, children exhibit the capacity to recall what was pretended a year *before* they can report their own earlier false beliefs. In Gopnik & Slaughter's experiment, all the children (3 and 4 year olds) performed at ceiling on remembering past pretenses, but most of the younger children (3 year olds) were unable to attribute false beliefs to themselves. A theory of pretense should be able to explain how the pretender's cognitive system succeeds in keeping what is really believed separate from what is pretended. It should also explain how the pretender can have accurate beliefs about what is being pretended.

While the contents of pretense episodes usually do not affect what the pretender really believes, pretense can often have a significant effect on other mental states. In particular, pretense can apparently have powerful emotional effects. In the CHILDES protocol, Abe seems to expect to be scared if they pretend that there is a monster coming. But there's little serious research on emotions in pretend play. However, as we'll explain in section 4, we think that our theory of pretense is also a theory of imagination. And there is plenty of research demonstrating that imagination can have a significant effect on emotions. Indeed, a standard way to elicit emotions in the laboratory is by asking the subjects to imagine something that will produce predictable affective responses (Izard 1991, 172). Imagining that something dangerous is happening can lead subjects to feel real fear, or something very like it. Imagining that one is having a heated conversation with an annoying acquaintance can lead to affect resembling real anger. Most of the research here depends on subjects' reporting their affect. However, there is a growing body of work showing physiological effects of imagining. Not surprisingly, there is evidence that imagining erotic encounters can lead to real sexual arousal, as evidenced by both self-report and physiological changes (e.g., Smith and Over 1987, 1990; Mosher et al. 1989). More recently, David Buss found that when males imagined their partner having sex with someone else, they showed marked physiological changes, including a significant increase in heart rate (Buss et al. 1992). On our view, the link between the imagination and affect raises lots of interesting questions to which neither we nor anyone else has answers. For instance, we don't know whether the imagination always has at

least weak links to the emotion system, no matter what the imagination is being used for; and we don't know whether the affective consequences of *imagining that p* always resemble the affective consequences of *believing that p*. We suspect that developing a detailed account of pretense will help clarify these sorts of issues.

4. A Theory About the Cognitive Mechanisms Underlying Pretense

4.1. Two framework assumptions.

In setting out our account of the cognitive mechanisms underlying pretense, we'll begin by sketching a pair of quite basic assumptions about the mind. Both assumptions are very familiar and we suspect that both of them are shared by most other people working in this area, though more often than not the assumptions are left tacit. We think it is important to be very explicit about them, since keeping the premises in mind forces us to be clear about many other details of our theory, details which other writers sometimes leave unspecified. The assumptions will serve as a framework upon which we will build as we develop our theory of pretense.

We'll call the first of our assumptions *the basic architecture assumption.* What it claims is that a well known commonsense account of the architecture of the cognitive mind is largely correct, though it is far from complete. This account of cognitive architecture, which has been widely adopted both in cognitive science and in philosophy, maintains that in normal humans, and probably in other organisms as well, the mind contains two quite different kinds of representational states, beliefs and desires. These two kinds of states differ "functionally" (as philosophers sometimes say) because they are caused in different ways and have different patterns of interaction with other components of the mind. Some beliefs are caused fairly directly by perception; others are derived from pre-existing beliefs via processes of deductive and non-deductive inference. Some desires (like the desire to get something to drink or the desire to get something to eat) are caused by systems that monitor various bodily states. Other desires, sometimes called "instrumental desires" or "sub-goals," are generated by a process of practical reasoning

that has access to beliefs and to pre-existing desires. The practical reasoning system must do more than merely generate sub-goals. It must also determine which structure of goals and sub-goals are to be acted upon at any time. Once made, that decision is passed on to various action controlling systems whose job it is to sequence and coordinate the behaviors necessary to carry out the decision. Figure 1 is a "boxological" rendition of the basic architecture assumption. We find diagrams like this to be very helpful in comparing and clarifying theories about mental mechanisms, and we'll make frequent use of them in this paper. It is important, however, that the diagrams not be misinterpreted. Positing a "box" in which a certain category of mental states are located is simply a way of depicting the fact that those states share an important cluster of causal properties that are not shared by other types of states in the system. There is no suggestion that all the states in the box share a spatial location in the brain. Nor does it follow that there can't be significant and systematic differences among the states within a box. When it becomes important to emphasize such differences, we use boxes within boxes or other obvious notational devices. All of this applies as well to processing mechanisms, like the inference mechanism and the practical reasoning mechanism, which we distinguish by using hexagonal boxes.

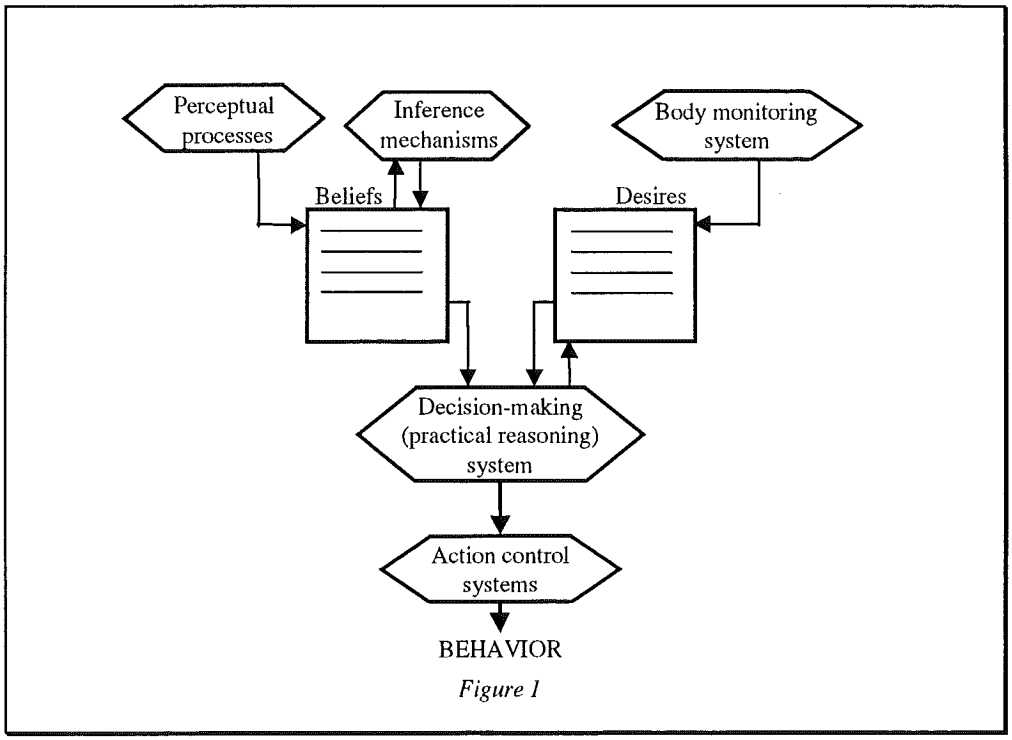

Figure 1

Our second assumption, which we'll call *the representational account of cognition*, maintains that beliefs, desires and other propositional attitudes are relational states. To have a belief or a desire with a particular content is to have a representation token with that content stored in the functionally appropriate way in the mind. So, for example, to believe that Socrates was an Athenian is to have a representation token whose content is *Socrates was an Athenian* stored in one's Belief Box, and to desire that it will be sunny tomorrow is to have a representation whose content is *It will be sunny tomorrow* stored in one's Desire Box.[5] Many advocates of the representational account of cognition also assume that the representation tokens subserving propositional attitudes are linguistic or quasi-linguistic in form. This additional assumption is no part of our theory, however. If it turns out that some propositional attitudes are subserved by representation tokens that are not plausibly viewed as having a quasi-linguistic structure, that's fine with us.[6]

4.2. The Possible World Box, the UpDater and the Script Elaborator: Three Further Hypotheses About Cognitive Architecture.

At the center of our theory of pretense are three further hypotheses about our cognitive architecture – three new "boxes" that we propose to add to the account depicted in Figure 1. The first of these is what we'll call *The Possible World Box (*or *the PWB).* Like the Belief Box and the Desire Box, the Possible World Box contains representation tokens. However, the functional role of these tokens – their pattern of interaction with other components of the mind – is quite different from the functional role of either beliefs or desires. Their job is not to represent the world as it is or as we'd like it to be, but rather to represent what the world would be like given some set of assumptions that we may neither believe to be true nor want to be true. The PWB is a work space in which our cognitive system builds and temporarily stores representations of one or another

[5] We will use italicized sentences to indicate representations or contents. Typically, the context will make clear whether we're referring to a content or to a representation.

[6] For more on the representational account of cognition see Sterelny (1990).

possible world.[7] We are inclined to think that the mind uses the PWB for a variety of tasks including conditional planning, reasoning about hypothetical situations and mind reading, and that dealing with one or more of these tasks may well have been the evolutionary function of the PWB. But in our theory the PWB also plays a central role in pretense. It is the workspace in which the representations that specify what is going on in a pretense episode are housed.

Early on in a typical episode of pretense, our theory maintains, one or more initial pretense premises are placed in the PWB workspace. So, for example, as a first approximation we might suppose that in the banana / telephone scenario the pretense initiating representation would be one with the content: *This* [banana] *is a telephone.* In Leslie's tea party pretense, it will be a representation with the content: *We are going to have a tea party.* What happens next is that the cognitive system starts to fill the PWB with an increasingly detailed description of what the world would be like if the initiating representation were true. In the banana / telephone case, the PWB presumably comes to contain representations with contents like the following: *People can use this thing* [the banana] *to talk to other people who are far away. In order to talk to people using this thing, you have to put one end near your ear and the other near your mouth. When someone wants to talk to you on this thing, it goes "ring, ring."* etc. In Leslie's tea party scenario, at the point in the pretense where *Alan has just turned the green cup upside down.* has been added to the PWB, the child's cognitive system has to arrange to get *The green cup is empty.* in there too.

How does this happen? How does the pretender's cognitive system manage to fill the PWB with representations that specify what is going on in the pretense episode? One important part of the story, on our theory, is that the inference mechanism, *the very same one that is used in the formation of real beliefs*, can work on representations in the PWB

---

[7] We are using the term "possible world" more broadly than it is often used in philosophy (e.g., Lewis 1986), because we want to be able to include descriptions of worlds that many would consider *impossible.* For instance, we want to allow that the possible world box can contain a representation with the content *There is a greatest prime number.*

in much the same way that it can work on representations in the Belief Box. In the course of a pretense episode, new representations get added to the PWB by *inferring* them from representations that are already there. But, of course, this process of inference is not going to get very far if the only thing that is in the PWB is the pretense initiating representation. From *This* [banana] *is a telephone.* there aren't many interesting inferences to be drawn. In order to fill out a rich and useful description of what the world would be like if the pretense-initiating representation were true, the system is going to require lots of additional information. Where is this information going to come from? The obvious answer, we think, is that the additional information is going to come from the pretender's Belief Box. So, as a first pass, let us assume that the inference mechanism elaborates a rich description of what the pretend world would be like by taking both the pretense-initiating representations *and* all the representations in the Belief Box as premises. Or, what amounts to the same thing, let us assume that in addition to the pretense initiating premise, the cognitive system puts the entire contents of the Belief Box into the Possible World Box.

There is, however, an obvious problem with this proposal. As we have told the story, when the inference mechanism is elaborating the pretend world description in the PWB it gets to look at what has been placed in the PWB and at *everything* in the Belief Box. This clearly can't be right, since it will typically be the case that one or more of the representations in the PWB is incompatible with something in the Belief Box. The pretender believes that the cup is empty (not full), that the banana is a banana (not a telephone), that he is a live person (not a dead cat) etc. So if the inference mechanism can look at *everything* in the Belief Box, it's going to generate glaring contradictions within the possible world description that's being built up in the Possible World Box. This would produce inferential chaos in the Possible World Box. Obviously this does not happen, as evidenced by the inferential orderliness displayed in the examples of pretense that we've collected. How can the theory handle this problem?

The answer, we think, is implicit in the fragment of our theory that we've already sketched. To see it, however, we have to step back and think about the operation of the

cognitive system while it is carrying out its normal, non-pretense, chores. One of the things that happens all the time is that via perception or via inference or from the report of another person, a cognitive agent learns a new fact or acquires a new belief that is incompatible with what he currently believes or with something entailed by what he currently believes. Nichols believes that his baby is fast asleep in her crib with her Teddy Bear at her side, but suddenly he hears the characteristic thump of Teddy hitting the floor, followed by giggling and cooing in the baby's room. It's a perfectly ordinary event which requires that his cognitive system update a few of his beliefs. Other cases are more dramatic and complicated. One day a long time ago Stich walked into his doctor's office and saw him listening attentively to his radio. "President Kennedy has been killed," he said. Suddenly a large number of Stich's beliefs had to be changed. How do our cognitive systems accomplish these tasks? It is notoriously the case that no one has been able to offer anything that even approximates a detailed account of how this process works. To provide such an account it would be necessary to explain how our cognitive systems distinguish those beliefs that need to be modified in the light of a newly acquired belief from those that do not. And to explain how we do that would be to solve the "frame problem" which has bedeviled cognitive science for decades (see, e.g., the essays in Pylyshyn 1987). Though we don't have any idea how the process of belief updating works, it is obvious that it *does* work and that it generally happens swiftly, reasonably accurately, and largely unconsciously. So there must be a cognitive mechanism (or a cluster of them) that subserves this process. We will call this mechanism the *UpDater*. And since the UpDater is required for the smooth operation of everyday cognition, it looks like we have reason to add another box to our boxological sketch of mental architecture. Some theorists might wish to portray the UpDater as a separate processing mechanism. But we are inclined to think it is best viewed as a sub-system in the inference mechanism, as indicated in Figure 2.

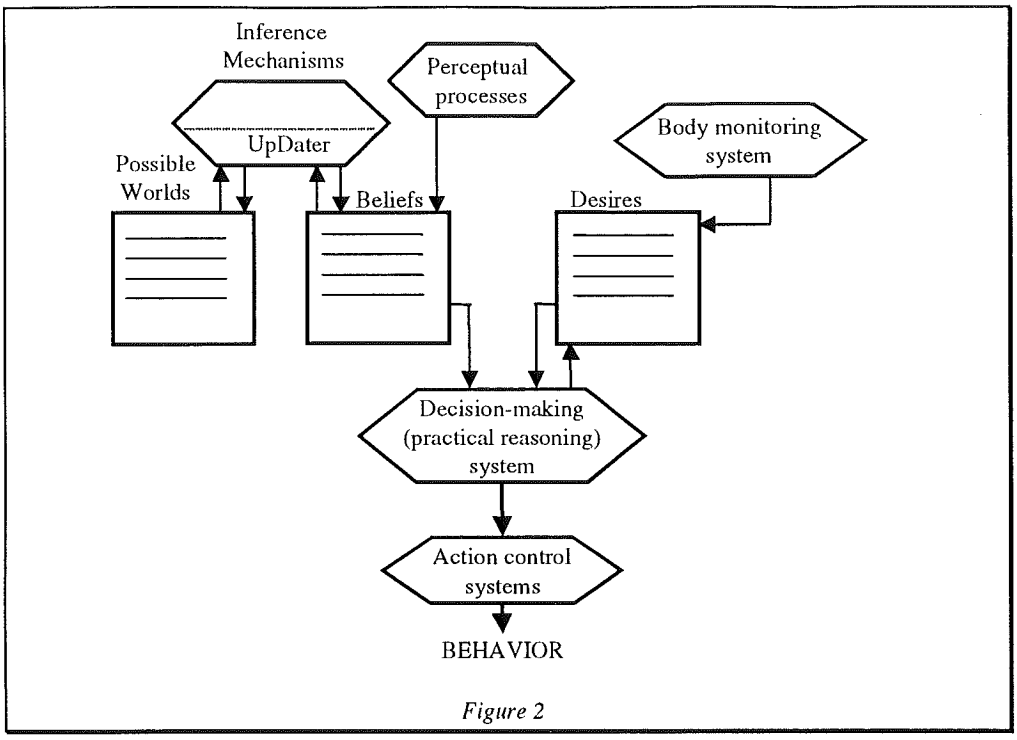

Figure 2

We have already assumed that the inference mechanism which is used in the formation of real beliefs can also work on representations in the PWB. Since the UpDater is a subcomponent of the inference mechanism, it too can work on the representations in the PWB. And this looks to be the obvious way of avoiding the explosion of contradictions that might otherwise arise when the pretense premises and the contents of the pretender's Belief Box are combined in the PWB. The basic idea is that when the pretense is initiated, the UpDater is called into service. It treats the contents of the Possible World Box in much the same way that it would treat the contents of the Belief Box when a new belief is added, though in the PWB it is the pretense premise that plays the role of the new belief. The UpDater goes through the representations in the PWB eliminating or changing those that are incompatible with the pretense premises. Thus these representations are unavailable as premises when the inference mechanism engages in inferential elaboration on the pretense premises. Alternatively, one might think of the UpDater as serving as a filter on what is allowed into the Possible World Box. Everything in the pretender's store of beliefs gets thrown into the possible world

box *except if it has been filtered out (i.e. altered or eliminated) by the UpDater.*[8] Obviously the UpDater will have lots to do during pretense since it is often the case that a large number of beliefs will have to be filtered out very quickly. But we don't think this counts as an objection to our theory. For the task the UpDater confronts in pretense is no more daunting than the tasks it must regularly handle in updating the Belief Box. There too it will often have to make lots of changes and make them very quickly.

We have suggested that the UpDater and other inference mechanisms treat the pretense representations in roughly the same way that the mechanisms treat real beliefs, but we have said little about the representational format and the logical form of pretense representations. One possibility that we find attractive is that the representations in the PWB are in the same "code" as representations in the Belief Box. The idea is that, although pretense representations are functionally distinguished from beliefs, there is an important sense in which they are similar. First, whatever determines the representational (or semantic) properties of representations in the Belief Box also determines the representational properties of representations in the PWB, and it does this in the same way in both cases. More importantly for our purposes, if pretense representations and beliefs are in the same code, then any processing mechanisms that work on the representations in the Belief Box and in the PWB will treat tokens of the two in the same way. We have already noted that we think that the inference mechanisms process pretense representations and beliefs in the same way. We suspect that there are other mechanisms, e.g., emotional systems, that also work on representations of both the Belief Box and the PWB. If there are such mechanisms, and if pretense representations and beliefs are in the same code, the mechanisms will process pretense representations in the same way the mechanisms process beliefs.

---

[8] The proposal can also be characterized in a slightly different way. Instead of claiming that all of the consistent beliefs are dumped into the PWB, we might say that the UpDater determines which beliefs are *relevant to* and consistent with the pretense premise, and only those beliefs get inserted into the PWB. On this characterization, there would be lots of beliefs in the Belief Box that are consistent with the pretense premise but are not part of the description being built up in the PWB because the beliefs are irrelevant to the pretense.

If pretense representations are in the same code as beliefs, then at least as far as the inference mechanisms are concerned, the pretense representation *Hamlet is the Prince of Denmark* has the same logical form as the belief representation *Charles is the Prince of Wales*. The issue of logical form becomes considerably more subtle and complex for pretense involving props, e.g., pretending that a banana is a telephone. One simple account is that the logical form of such a pretense representation is just *This banana is a telephone*., analogous to the logical form of the belief representation, *This animal is a mammal*. However, some of the responses in our studies speak against this simple account. For a number of our participants, we asked several questions about their pretense while they were pretending that the banana was a telephone. For example, we asked whether they were pretending that you could sometimes eat a telephone or whether they were pretending that you could sometimes peel a telephone. The subjects claimed that they weren't pretending that you could sometimes eat or peel a telephone. This suggests that subjects don't include *banana*, with all its attendant properties, as part of the pretense representation. As an alternative to the simple account, we are attracted to the idea that demonstratives often play a crucial role in pretense representations involving props. On this account, in pretending that a banana is a telephone, one has the pretense representation, *This is a telephone*, where "*this*" refers to the banana. We have tried to capture this by putting the referent in brackets, e.g., *This* [banana] *is a telephone.* This 'demonstrative' account can accommodate the responses in our studies. Since the demonstrative ("*this*") need not be tied to any particular properties of the prop, there is no reason to expect the pretender to infer that her pretend telephone is edible or peelable. Another advantage to construing such pretense representations as implicating demonstratives is that this fits with the idea that pretense representations have the same logical form as belief representations. It is plausible that there are "demonstrative thoughts", e.g., *THAT's going to explode*, or *THIS is bigger than THAT*. Indeed, there is a substantial literature about indexicals and demonstrative thoughts in philosophy and cognitive science (e.g., Peacock 1983, Perry 1993, Leslie et al. 1998). Perhaps, then, the UpDater and other inference mechanisms treat 'demonstrative pretenses' in much the same way the mechanisms treat 'demonstrative thoughts'.[9]

---

[9]We do not mean to imply that the features of the prop are *never* relevant to the pretense.

In claiming that the UpDater treats the contents of the PWB in much the same way that it treats the contents of the Belief Box, we want to leave open the possibility that there may be some systematic differences to found. There is some intriguing evidence suggesting that emotional and motivational factors may affect either the thoroughness with which the UpDater goes about its work in the Belief Box, or the standards it applies in determining whether new evidence is strong enough to lead to the elimination of an old belief, or both. For instance, Ziva Kunda (1987) argues that motivational factors produce self-serving biases in inference. In one of her experiments, Kunda presented subjects with putative evidence on the negative effects of caffeine consumption. She found that heavy caffeine users were much less likely to believe the evidence than low caffeine users (Kunda 1987). It might well be the case that motivational factors play an important role when the UpDater is working on the contents of the Belief Box but that motivational factors play much less of a role when the UpDater is working on the contents of the Possible World Box. It is, we think, a virtue of our strategy of boxological explicitness that it brings empirical issues like this into much sharper focus.

We assume that the contents of a pretender's Belief Box include not only representations whose contents are individual propositions – like the belief that bananas are yellow – but also clusters or packets of representations whose contents constitute "scripts" or "paradigms" detailing the way in which certain situations typically unfold (see e.g., Schank and Abelson 1977, Abelson 1981). These scripts often play an important role in guiding and constraining the description of a possible world which gets elaborated in the course of a pretense episode.[10] So, for example, all the subjects who

---

Paul Harris notes that children are more likely to use a banana than an orange as a pretend telephone (personal communication). This does not undermine the parallel with demonstrative thought since demonstrative thoughts can also be influenced by features of the object. For example, the thought *This is bigger than that.* is typically influenced by the size of the objects. In pretense, we can largely decide which features of the props will be relevant to the pretense; similarly, we can decide which prop would work best for the pretense.

[10] In a number of recent papers Paul Harris (1993, 1994) has emphasized the importance of scripts and paradigms in pretense and imagination. We are indebted to Harris for prompting us to think more carefully about these matters.

participated in our fast food restaurant scenario followed the standard pattern of the fast food restaurant script: Order first, then pay and get the food, then eat. But while scripts can provide the general structure for many pretense episodes, they leave many aspects of the episode unspecified. Some additional constraints are imposed by the details of what has gone on earlier in the pretense along with the pretender's background knowledge. In one run through of our fast food scenario, for example, the "server" decided to pretend that she was working at Wendy's and began her pretense by saying "Hi. Welcome to Wendy's. Can I take your order?" At this point the "customer's" background knowledge of the Wendy's menu served to constrain the sorts of things that she could pretend to be ordering. Hamburgers of various sorts were OK, as were French fries, but hot dogs, tacos and egg rolls were not. This still leaves many options open, however. Our subject decided to order a small hamburger with ketchup, jumbo fries and a large sweet tea. In the post-pretense interview she explained that she always had hamburgers with ketchup, though she was aware that there were a variety of other things that she might have ordered instead. So, within the script-constraints, there are a variety of choices that the pretender has to make. In the protocols that we collected, sometimes the pretender's choices followed what the pretender would normally do in real life. Other times, however, the pretender deviated from what he would normally do. In one of our (unreported) fancy restaurant episodes, the "diner" behaved in a boorish manner, but after the pretense he claimed that he would never really act that way in a restaurant. When asked why he chose to be a surly customer, the participant said, "Because playing is more fun. It's fun to not be yourself." In addition to the pretender's decisions about what she herself will do, sometimes the pretender must develop the pretense by deciding what happens next in the pretended environment: Does the banana / telephone ring? If so, who is calling? What does the caller want?

Although the pretender's choices often conform to script-constraints, the script-constraints themselves are only "soft" constraints, since a really imaginative pretender might order sushi at Wendy's just to make the little skit more interesting. Indeed, some of our subjects violated the script-constraints rather dramatically. For example, in the second fancy restaurant scenario we reported, the waiter pretended to crush peppercorns

with the heel of his shoe, he gave the diner a sword to cut lamb chops, and he killed one of the patrons with the sword. In these examples, the pretender's decision about what to do goes well beyond the restaurant script.

The point of all of this is to emphasize that pretense is full of choices that are not dictated by the pretense premise, or by the scripts and background knowledge that the pretender brings to the pretense episode. The fact that these choices typically get made quite effortlessly requires an explanation, of course. And we don't pretend to have a detailed account of the cognitive mechanisms that underlie this process. There must, however, be *some* mechanism (or, more likely, a cluster of mechanisms) that subserves this process of script elaboration. So we propose to add yet another item to our boxology, the *Script Elaborator* whose job it is to fill in those details of a pretense that can't be inferred from the pretense premise, the (filtered) contents of the Belief Box and the pretender's knowledge of what has happened earlier on in the pretense.

4.3. Explaining Pretense Behavior: Information and Motivation

Figure 3 is a boxological rendition of the cognitive mechanisms that we now have in place in our theory.

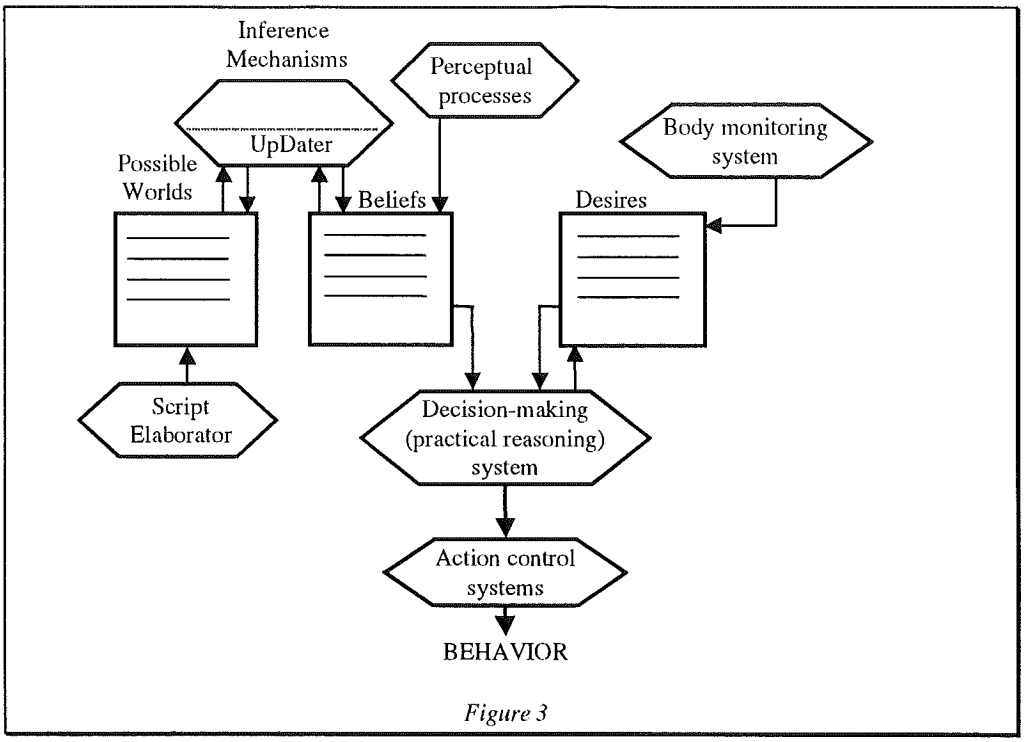

Figure 3

Those mechanisms provide at least the beginnings of an explanation for several of the features of pretense set out in Section 3. Since the representations which describe what is going on in the pretense episode are stored in the Possible World Box, not in the Belief Box, we have a ready explanation for the cognitive quarantine that is characteristic of pretense. People who pretend that there is a burglar in the basement don't end up believing that there is a burglar in the basement, because they never put a representation with the content *There is a burglar in the basement* in their Belief Box. Similarly, a child who pretends that a banana is a telephone never has a representation of *This* [banana] *is a telephone* in his Belief Box, and thus there is no danger that the child will come to believe that some bananas really are telephones. Another part of our theory claims that the inference mechanism, which adds, modifies and removes representations in the Belief Box, also adds, modified and removes representations in the PWB. This, along with our hypothesis that the pretense premise and most of the contents of the Belief Box are available in the PWB at the start of an inference episode, provides a framework for explaining the sort of inferential elaboration that occurs in pretense. There is, however, one quite crucial aspect of pretense for which our theory has not yet provided any explanation at all. It does not explain why pretenders *do* anything; it offers no

explanation of their *behavior*. Rather, what the theory explains is how a cognitive agent can go about conceiving of or imagining a world which is different from the actual world. So, while it might be offered as a theory of imagination (and, indeed, we maintain that it is a plausible theory of imagination) it is not yet a theory that is able to explain pretense.

Why does a person who is engaging in pretense do the sometimes very peculiar things that pretenders do? Why, for example, does a child or an adult who is pretending to be a train walk around making jerky movements and saying "Chugga Chugga, Choo Choo?" The answer we would propose comes in two parts, the first of which is really quite simple. Pretenders behave the way they do because they *want to behave in a way that is similar to the way some character or object behaves in the possible world whose description is contained in the Possible World Box.* To pretend that p is (at least to a rough first approximation) to behave in a way that is similar to the way one would (or might) behave if p were the case. (See Lillard 1994, 213 for a similar treatment.) Thus a person who wants to pretend that p wants to behave more or less as he would if p were the case. In order to fulfill this desire, of course, the pretender must know (or at least have some beliefs about) how he would behave if p were the case. And the obvious source of this information is the possible world description unfolding in the PWB. However, since the PWB is distinct from the Belief Box, we must assume that the contents of the former are accessible to the latter. More specifically (and this is the second part of our answer) we assume that as a possible world description is unfolding in the PWB, the pretender comes to have beliefs of the form: If it were the case that p, then it would (or might) be the case that q_1 & q_2 & ... & q_n, where p is the pretense premise and q_1 & q_2 & ... & q_n are the representations in the PWB. These beliefs, along with the desire to pretend, lead to the pretense behavior in much the same way that Stich's belief that Nichols has just walked around making jerky motions and saying "Chugga Chugga, Choo Choo" and Stich's desire to behave in a way that is similar to the way in which Nichols behaved will lead Stich to walk around making jerky motions and saying "Chugga Chugga, Choo Choo."

Although these beliefs concerning conditionals derive from the PWB, such beliefs should not be regarded as beliefs about pretense. As we'll explain in section 5, we think that it's possible for young children to pretend without having any beliefs about pretense. In those cases, the child might have a conditional belief that guides the pretend behavior, but not any beliefs about pretense. Nonetheless, adults and older children clearly do have beliefs about what they are pretending, and they can report on those beliefs. For instance, in our studies, we interrupted a participant during the cake-baking scenario to ask, "What are you doing now?". The participant said, "I'm pretending that I'm stirring the egg and the flour and the milk." Obviously there must be some set of mechanisms that enable people to recognize and report their own pretenses. This implicates difficult and controversial issues about self-awareness (e.g., Gopnik 1993; Goldman 1993), and we want to skirt those issues as much as possible. For our purposes, it suffices to note that *somehow* we are able to report our own beliefs and desires. However it is that we recognize and report on our own beliefs and desires -- whether it's mediated by a theory of mind, utilizes mechanisms, or (as we suspect) some combination of the two -- we might exploit the same (or similar) mechanisms to recognize and report on our pretenses (for more details, see Nichols & Stich forthcoming).

It is worth emphasizing that the pretense initiating desire – the desire to behave in a way *similar* to the way in which some character or object behaves in the possible world described in the PWB – is typically not a desire to behave in *exactly* the same way. Just how close an approximation the behavior will be will depend on many factors, including the pretenders other desires and his knowledge about the consequences of various actions. Thus, for example, in our burglar in the basement scenario, one subject picked up the phone that was available and dialed 911. However, she took precautions to insure that the call did not really go through. She didn't want her behavior to be *that* similar to the behavior elaborated in the PWB; she wanted to be sure that the police didn't really come.

We have suggested that the pretender engages in the pretense action because she *wants* to behave in a way similar to the way that some person or object would behave in a

possible world scenario. But there is another motivation question that we haven't addressed. We don't explain why people would want to engage in pretend play at all. Our suspicion is that this question, like the question, "why do people want to have sex?", may require an evolutionary explanation rather than an explanation in terms of further motivational and cognitive features.[11]

Obviously, what we have presented in this section is, at best, just the bare bones of a theory of pretense. There are lots of details that we have left unspecified. Despite this, however, we maintain that our theory provides a more promising framework for explaining the facts of pretense than any of the other accounts that have been offered. It is also, in many quite crucial ways, much *more* detailed than other accounts to be found in the literature. In the section that follows, we'll defend this view by comparing our theory with the competition.

5. A Comparison With Other Theories

One of our central themes in this section is that theories about the cognitive mechanisms underlying pretense that have been offered by other authors are seriously incomplete. They simply do not tell us how the theory would explain many of the most salient features of pretense, features like those that we have assembled in Section 3. A second central theme is that, when one sets out to elaborate and amend these alternative

---

[11] Carruthers (1996) offers a story about why people engage in pretend play. He suggests that "children find rewarding that feature which is common to *all* forms of pretend-play, namely the manipulation of the child's own mental states, through supposing or imagining", and on Carruthers' view, "you cannot enjoy supposing or imagining without being conscious of your (mental) activity. In general, *enjoying Xing* presupposes *awareness of Xing*, – which is why you cannot enjoy digestion, sleepwalking, or subliminal perception" (265). We are not particularly sanguine about this account. Even if enjoyment does require awareness, surely conscious enjoyment is not the only kind of motivational state, and it's not clear why we should assume that the motivation for pretend play is always based on conscious enjoyment. Indeed, Carruthers' motivational story has the awkward consequence that very young children must be aware of their own mental states of pretending (265-6). Furthermore, even if Carruthers' story were right, it wouldn't explain pretend *play*, since Carruthers' story says nothing about why children would want to carry out the scenarios they construct in imagination.

accounts to enable them to explain the facts that they cannot otherwise explain, the most promising proposals tend to make the competing theories look a lot like ours. If, as we would hope, other authors view our suggestions as friendly amendments to their theories, it may well be the case that something like the framework we have presented will emerge as a consensus toward which many theorists are heading from many different directions.

Though there are many suggestions about the cognitive processes underlying pretense to be found in the literature, we think that for the most part, the accounts fall into two distinct clusters. The central idea of one of these clusters is that pretense is subserved by a process of simulation which is quite similar to the off-line simulation process that, according to some theorists, underlies our capacity to predict people's decisions and other mental states. For reasons that will emerge shortly, we will call these "on-line simulation" accounts. The central idea of the second cluster is that pretense is subserved by a special sort of representational state, a "metarepresentation". We will consider on-line simulation accounts in Section 5.1 and metarepresentational accounts in Section 5.2. In Section 5.3, we will consider one alternative that does not fit easily into either category.

5.1. On-Line Simulation Accounts of Pretense

The off-line simulation account of mental state prediction was originally proposed to explain how we predict the behavior of someone whose beliefs or desires are different from our own. How, for example, might Stich go about predicting what Nichols would do if Nichols were at home alone at night and heard sounds that led him to believe there was a burglar in the basement? On the off-line simulation account, the prediction process proceeds as follows. First, Stich (or, more accurately, some component of his cognitive system) adds a special sort of belief (often called an "imaginary" or "pretend" belief) to his pre-existing store of beliefs. This "imaginary" belief would have a content similar or identical to the content of the belief that Nichols would actually have in the situation in question. For purposes of the illustration, we can suppose that the imaginary belief has the content *there is a burglar in the basement.* In many crucial respects imaginary beliefs

have the same causal powers as real ones. Thus once the imaginary belief is added to Stich's Belief Box, his cognitive system sets about doing many of the things that it would *actually* do if Stich really believed that there was a burglar in the basement. The result, let us assume, is a decision to reach for the phone and dial 9-1-1 in order to summon the police. However, one of the ways in which decisions that result from imaginary beliefs differ from decisions that result from real beliefs is that the cognitive agent does not really *act* on them. Rather, the decision that results from the imaginary belief is shunted "off-line" to a special cognitive mechanism which embeds the content of the decision in a belief about what the "target" (Nichols) will decide. In this case the belief that is formed is that Nichols will decide to reach for the phone and dial 9-1-1. Figure 4 is a boxological sketch of this process.

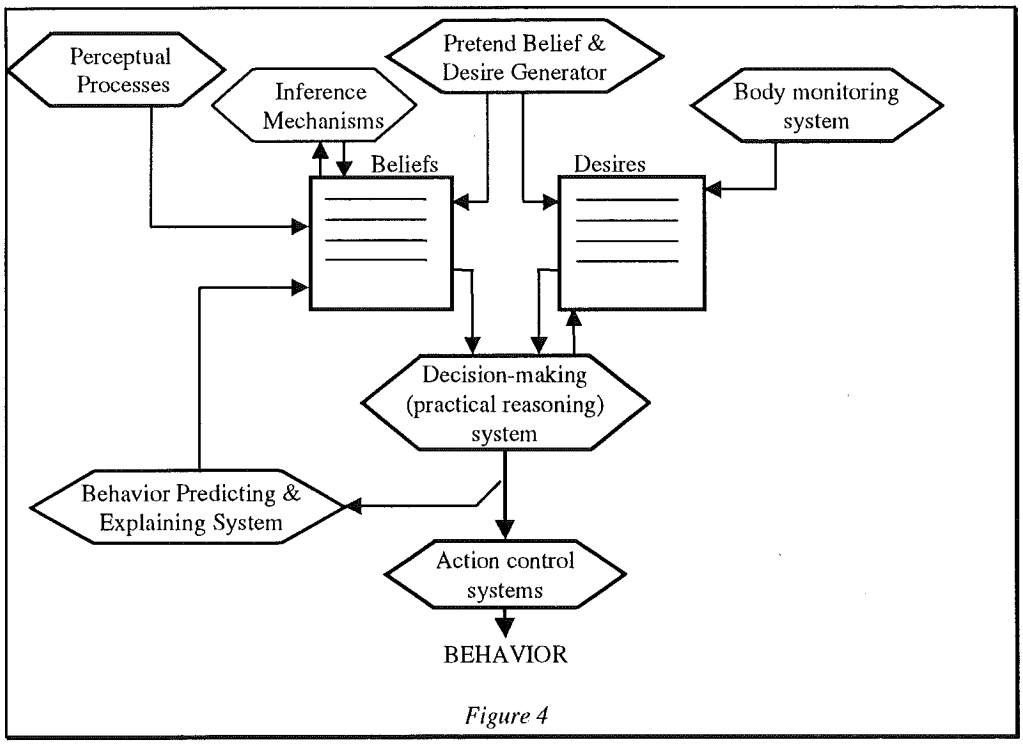

Figure 4

A number of theorists who accept this account of how we go about predicting people's decisions have suggested that, with a few modifications, it might also serve as an account of the mental processes subserving pretense. The first modification is that in pretense the imaginary belief that is added to the Belief Box is not a belief attributed to

some target whose behavior we want to predict. Rather it will be what we earlier called an "initial pretense premise" (or several such premises) whose content specifies the basic assumption of an impending pretense episode. So, for example, if Stich chose to *pretend* that there was a burglar in the basement, the episode would start with an imaginary belief with the content *there is a burglar in the basement* being placed in his Belief Box. As in the case of decision prediction, once the imaginary belief is added to the Belief Box, the cognitive system sets about doing many of the things that it would *actually* do if the pretender believed that there was a burglar in the basement. And, as before, we will assume that the result of this process, in Stich's case, would be a decision to reach for the phone and dial 9-1-1 to summon the police. In the pretense case, however, the decision is not taken "off-line". Rather, Stich actually does reach for the phone and dial 9-1-1. So pretense, on this account, is very much like off-line prediction of people's decisions, except that the imagination driven decision is not taken off line. The pretender actually carries it out. Figure 5 is a boxological rendition of this "on-line simulation" account of pretense.

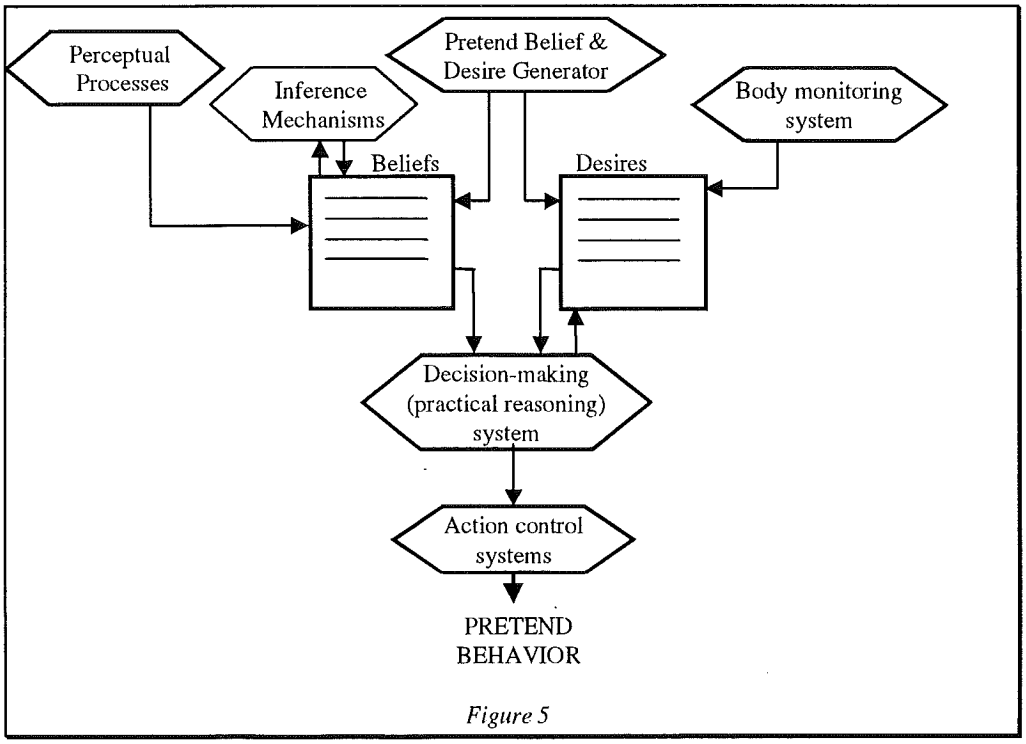

Figure 5

Robert Gordon has been a leader in developing the off-line simulation account of mental state prediction, and though his discussion of pretense is quite brief and sketchy, we are inclined to think the account we have just set out is a plausible interpretation of the theory of pretense proposed by Gordon in collaboration with John Barker. "In pretense," they write

> [children] accept an initial premise (or premises) -- for example, that certain gobs of mud are pies. By combining the initially stipulated premise with their existing store of beliefs and calling upon their reasoning capacity, they are able to obtain answers to questions not addressed in the initial premise. (Gordon & Barker 1994, 171)

Two paragraphs later, Gordon & Barker sketch an off-line simulation account of decision and behavior prediction, and explicitly draw the parallel with pretense:

> As in pretend play, an initial premise -- here, the hypothetical condition -- is added to one's store of beliefs, desires, and other inputs to intention-formation and decision making. In one important respect, however, this kind of simulation is unlike children's games of make believe .... It stops short of overt action. (172)

Unfortunately, Gordon and Barker do not say much more about the processes subserving pretense, and thus we can't be certain that Figure 5 really captures what they have in mind. We are, however, much more certain that if this *is* what they have in mind, then the theory confronts some quite fundamental problems.

5.1.1. Some Problems for the Gordon & Barker Theory of Pretense

The first, and most obvious of these problems is that, as we have interpreted it, the Gordon and Barker theory offers no way of explaining the phenomenon of cognitive quarantine. If, as Gordon and Barker suggest, the pretense initiating "hypothetical condition" really is simply "*added* to one's store of beliefs, desires and other inputs to intention formation" then, it would seem, the pretender will actually believe the premise

and anything inferred from it in the course of the pretense. Moreover, when the episode of pretense is over, the pretense premise and everything inferred from it will still be sitting around in the pretender's Belief Box; Gordon and Barker give us no hint about how they propose to get them out. So, in burglar in the basement example, Stich should actually believe that there is a burglar in the basement, that an intruder might at any moment appear at the top of the stairs, that he is in serious danger, and so on. Moreover, these beliefs should remain in place after the episode of pretense has ended. Obviously nothing like that occurs in typical cases of pretense. So, at the very least, the sort of theory we're attributing to Gordon and Barker will have to add some bells and whistles to explain why all this *doesn't* happen.

A second problem with the Gordon and Barker theory is that it offers no explanation for the fact that when pretense assumptions are added to the pretender's store of beliefs, and the inference mechanism does its work, the result is not simply a chaotic stream of contradictions. When Stich pretends that there is a burglar in the basement he simultaneously believes that there is no one in the basement. (If he didn't believe that he'd stop pretending in a big hurry. There would be more important things to do.) So it would appear that on Gordon and Barker's account Stich has two representations in his Belief Box, one with the content *There is a burglar in the basement* and one with the content *There is no one in the basement.* Something must be said to explain why these patently incompatible beliefs don't lead to an inferential meltdown.

A third problem confronting the sort of theory sketched in Figure 5 is that in many cases it provides no plausible account of the pretender's motivation to perform the actions he performs in the course of the pretense. The problem does not arise in the burglar example, since if Stich really did believe there was a burglar in the basement, he would certainly want to summon the police. Though one might fuss about the details, the rough etiology of that desire are clear enough: Stich believes that if there is an intruder in his home there is a serious risk that he will be injured. He also believes that one good way to prevent injury under those circumstances is to summon the police. And, of course, he has a long standing desire not to be injured. These background beliefs and that

long standing desire would play a central role in generating a decision to summon the police, if Stich actually believed that there was a intruder in his home. So if the pretense premise that there is a burglar in the basement and everything inferred from it have causal powers similar to the real belief that there is a burglar in the basement, then Stich's motivation in reaching for the phone is not at all puzzling. But now let us shift to another case: the child making mud pies. Here, according to Gordon and Barker, the pretense premise is that certain gobs of mud are pies. From this, let us suppose, the pretender infers that the gobs of mud can be eaten. But why does he actually pick them up and (pretend to) eat them? There are, after all, lots of occasions on which a child sees a *real* pie, but does not try to eat it because he's simply not hungry. And it's clearly possible for a child (or an adult) to engage in the mud pie pretense even though he is not at all hungry. Suppose this is the case in the example at hand. The child has just had a substantial meal and has no desire at all to eat anything else. Thus there is no current desire to play the motivating role that the long standing desire not to be injured plays in the burglar example. So what explains the pretender's behavior? Why does the imaginary belief that the gobs of mud are pies lead to (pretend) eating? The theory that we've attributed to Gordon and Barker offers no answer. As we'll soon see, there are a variety of ways in which this theory might be amended to handle the problem. In deciding among them, there is one class of cases in which the problem of the pretender's motivation proves to be quite critical. These are the cases in which the pretender is pretending to be something that has no desires or motivations of its own. As we noted earlier, neither children nor adults find it in the least challenging to pretend to be a train or a dead cat. And, obviously, trains just aren't the sorts of things that have desires, and neither are dead cats. Moreover, while young children might have some mistaken beliefs about the desires of trains and dead cats, our adult subjects certainly did not.

The fourth and final problem on our list is that there are often quite predictable gaps between what a person would do if she believed that the pretense premise were true and what she does when pretending that it is true. Children who are pretending that mud pies are real pies typically don't really eat the mud pies, even if they happen to be hungry. They just make motions that roughly approximate those they would make if they

were really eating. Similarly, in the burglar in the basement pretense, the pretender doesn't really summon the police. Rather, he behaves in a way similar to the way that he would if he were really calling the police, taking care that the call does not actually go through. (Only one of our adult pretense subjects opted to call the police in the burglar-in-the-basement scenario. She reached for the phone and touched 9-1-1 without actually pressing the buttons.) On the Gordon and Barker account, these differences between real and pretend behavior present us with a puzzle. If an imaginary belief has causal powers similar to those of a real belief, and if the real belief would lead a subject to call the police, then why doesn't the imaginary belief have the same result? The theory we're attributing to Gordon and Barker offers no explanation.[12]

5.1.2. Some Proposed Extensions to the Gordon & Barker Theory of Pretense

In this section we propose to explore some of the ways in which Gordon and Barker's theory of pretense might be extended and modified to handle the problems pointed out in the previous section. As we noted earlier, it's our contention that the most promising proposals for patching the Gordon and Barker theory transform their theory into one that is indistinguishable from ours.

The first problem we posed for the Gordon and Barker theory was that it offered no explanation for the phenomenon of cognitive quarantine. If pretense premises are simply thrown into the Belief Box and treated like real beliefs, why don't pretenders end up believing both the premises and everything inferred from them? To handle the problem, perhaps the most obvious proposal is that, though pretense premises get added to the Belief Box, they must come specially marked in some way, and this marking insures that (i) they aren't treated as real beliefs except in the context of an episode of pretense, (ii) they don't get left behind after the pretense is over, and (iii) neither do any of the representations that are inferred from pretense premises during the course of the

[12] Gordon and Barker do note that "One does not carry out the decision, say to call the police, even in overt pretend play" (172). However, they don't explain how this fits with

pretense. But, of course, to say that the pretense premises and everything inferred from them have a special marker when thrown into the Belief Box, and that this special marker has important consequences for how the pretense-subserving-representations are treated, is tantamount to saying that these pretense-subserving-representations are functionally different from the other representations in the Belief Box. And since the "box" metaphor is just a way of distinguishing representations that have systematically different functional or computational properties, to say that pretense-subserving representations are functionally distinct from other representations in the Belief Box is equivalent to saying that they are in a box of their own. So the obvious way for Gordon and Barker to handle the problem of cognitive quarantine is to posit a Pretense Box which is similar to the Possible World Box posited in our theory. The Pretense Box is a functionally distinct component of the mind, a workplace in which pretense-subserving representations are stored and elaborated.

In the original version of the Gordon and Barker theory, in which pretense premises are simply added to the Belief Box, the inference mechanism that is used in ordinary, non-pretense cognition can also be used in drawing inferences within an episode of pretense. This idea, which is very much in the spirit of simulation theory, can be preserved by supposing that the inference mechanism can work in much the same way on representations in the Belief Box and on representations in the Pretense Box. Another important feature of the original Gordon and Barker theory is that, in elaborating inferences within an episode of pretense, the inference mechanism can use not only the pretense premise(s) but also all the background knowledge stored in the Belief Box. The obvious way to retain this feature, once the Pretense Box has been added, is to suppose, much as we did in our theory, that in elaborating pretense inferences, the inference mechanism can use all the representations in the Belief Box as premises. (Or, what amounts to the same thing, we can simply suppose that all the representations in the Belief Box function as though they were also within the Pretense Box while an episode of pretense is unfolding.) But, of course, this assumption, generates a problem that parallels the second problem we noted with Gordon and Barker's original proposal: The pretense

---

their claim that in games of make believe, the pretend decisions result in overt action.

premise(s) will typically be inconsistent with many of the representations in the Belief Box, and when the inference mechanism sets to work on this inconsistent set of representations, the result will be inferential chaos. By now, no doubt, the reader will have anticipated the suggestion we would make to deal with this problem. Since there are independent reasons (set out in Section 4.2) to posit an UpDater mechanism whose job it is to make appropriate changes in the Belief Box when new representations are added, an amendment to the Gordon and Barker theory can stipulate that the UpDater filters and modifies the contents of the Belief Box for compatibility with the pretense premise before they are allowed in to the Pretense Box. We now have solutions to the first two problems posed in 5.1.1, and the revised theory is growing to look a lot like ours.

The third problem we posed for the original version of Gordon and Barker's theory was that, in many cases, their theory offered no explanation of the motivation that led pretenders to behave. If participants in the mud pie pretense didn't happen to be hungry, for example, the Gordon and Barker theory could not explain why an imaginary belief that the mud gobs were pies (however much it might be like a real belief) leads to pretend eating behavior. There is some textual evidence that Gordon and Barker and others who urge accounts similar to theirs would deal with this problem by positing that, in some instances of pretense, imaginary desires are added to the Desire Box in much the same way that imaginary beliefs are added to the Belief Box. So, in the mud pie pretense, the theory would propose that an imaginary desire with the content *I eat some pie.* gets added to the real desires in the pretender's Desire Box.[13] However, for two quite different reasons, we are inclined to think that this is not the best way to deal with the problem. The first reason is that if this imaginary desire has causal powers similar to the causal powers of the real desire with the same content, then the desire would produce the wrong behavior. An (imaginary) desire to eat some pie along with an (imaginary)

[13] The passage we quoted from Gordon & Barker in section 5.1 suggests that on their view pretend desires do get added to the Desire Box. Gordon & Barker write, "As in pretend play, an initial premise... is added to one's store of beliefs, *desires,* and other inputs to intention formation and decision making" (172, our emphasis). This aspect of a simulationist account is more explicit in Currie, whose simulation-based view is

belief that the gob of mud is a pie would presumably lead the pretender to actually eat the mud pies. But pretense behavior of this sort is rarely seen in children and almost never seen in adults. The second reason is that there are cases in which the imaginary desire account looks to be quite hopeless. Though it is conceivable that an adult who is pretending to be a dead cat has added an imaginary representation with the content: *I am a dead cat.* to her Belief or Pretense Box, there is no plausible candidate for the imaginary desire which might interact with this belief to produce the sort of behavior that we described in Section 2. Dead cats don't have desires. A much better way to patch the Gordon and Barker theory, we maintain, is to drop the idea of imaginary desires altogether and to explain a pretender's motivation the way we did in Section 4.3. What is in the Pretense Box is not a pretend belief that might be attributed to a dead cat. Rather, the Pretense Box has a representation with the content: *This* [my body] *is a dead cat*, along with some additional representations about what the dead cat looks like. And what leads the pretender to do what she does is not some bizarre desire that might be attributed to a dead cat, but simply a quite real desire to "behave" in a way that is similar to the way a dead cat described in the Pretense Box would behave. In the mud pie case, what's in the Pretense Box might be a set of representations describing what the world would be like if the mud gobs were pies and someone was eating them. The motivating desire is not an imaginary desire to eat a pie, but a real desire to behave in a way that is similar to the way that the imagined person behaves.[14] But, of course, the pretender also has a competing desire not to eat mud, so she does not want to behave in exactly the way that a person fitting the description would behave, since the pretender knows that if she did she would get a mouth full of mud.

---

discussed in the next section.

[14] At least for older children and adults, the description being built up in the Pretend Box might include representations about desires or motivational states. Obviously, the Belief Box can contain representations about desires, e.g., the Belief Box might contain a representation with the content *I want some coffee.* Similarly, then, the Pretend Box might contain representations about desires, e.g., the Pretend Box might contain a representation with the content *I want some pie.* The motivating desire, though is not in the Pretend Box. Rather, the motivating desire is the real desire to behave in a way that fits the description being built up in the Pretend Box.

5.1.3. Other Simulation Based Accounts of Pretense

A number of other authors who are sympathetic to off-line simulation accounts of mental state prediction have suggested that the off-line simulation account of prediction can be extended to provide accounts of both pretense and imagination. In their 1993 monograph, Harris and Kavanaugh offer a "flagging" account of pretense according to which pretend episodes depend on "flagging" objects with make-believe stipulations (Harris & Kavanaugh 1993, 60). More recently, Harris revised the flagging theory to say that in pretend play, we "adopt a make-believe stance," by which he means that "the representation of the episode is prefaced by an attitudinal marker equivalent to the statement: 'In this episode, imagine that...'" (1995, 177) Though Harris does not tell us where these sorts of representation are located we don't think it's too much of a stretch to read him as positing something rather like the Pretense Box that we urged on Gordon and Barker. This interpretation is supported by his claim that in normal development children acquire "an increasing capacity for the internal creation of a complex script" which is detached from "the external context." (Harris 1993, p. 234) Since the representations in the child's Belief Box are closely tied to the external context, this "complex script" must be elaborated in some functionally distinct component of the mind. Harris (1991) has also stressed the importance of using one's own beliefs as "default settings" in imagination and pretense, and the role of stored prototypes in the elaboration of a pretense script. We have taken both of these ideas on board in our own theory. So we see Harris's account of pretense as broadly compatible with our own, though he is silent on a number of crucial issues including the avoidance of contradictions in pretense "scripts," the motivation for pretense behavior and the systematic differences between pretense and non-pretense behavior.

Gregory Currie's account of pretense is much sketchier than Harris's and much harder to interpret. According to Currie: "It is natural to think that childhood games of pretense in which the actors play at being pirates or bears are driven by the actors mentally taking on these roles and imagining being pirates or bears. In doing these things, the children simulate, to the best of their ability and within the compass of their

relevant knowledge, the experience of being pirates or bears" (Currie 1996, 251) Unfortunately, Currie tells us little about the mental mechanisms that subserve this sort of simulation and how they are related to the mental mechanisms involved in non-pretense cognition. In another passage in which he discusses "a *simulationist* theory of pretending" he refers the reader to Goldman's work on mental simulation, and Goldman has explicitly endorsed a model similar to Figure 4 as a plausible boxological rendition of his account of simulation (Goldman 1992). Currie also refers the reader to Gordon's work. So there is some reason to suppose that Currie might be interpreted as endorsing an account of pretense along the lines depicted in Figure 5. However, Currie also insists that pretending is a propositional attitude which is distinct from both believing and desiring, and that pretend beliefs have functional roles that are systematically different from real ones (Currie, forthcoming). So there is also some reason to suppose that Currie's account of pretense would include a separate Pretense Box. Indeed, the following passage suggests that Currie's account may also include two Pretense Boxes, one for pretend beliefs and one for pretend desires: "If mental simulation is how we understand the minds of others, there must be states that stand to desire as imaginings stand to beliefs; for us adequately to simulate the mental states of others, we must simulate their desires as well as their beliefs. So as well as pretend or imaginary beliefs – what I have been calling imaginings – we need pretend or imaginary desires" (Currie 1995a, 150). This, then, seems to be a point on which our theory and Currie's diverge. For the reasons set out in 5.1.2, we think that positing pretend desires raises more problems than it solves, and that our theory offers a more plausible account of the motivation of pretense behavior.

There is one aspect of Currie's theory that we think deserves special mention: His account of the evolutionary origin and biological function of pretense. According to Currie, the function of the mechanisms subserving imagination and pretense is to "test-run" possible courses of action. "Strategy testing," he writes, "is, on my hypothesis, the proper function of the imagination: the function appeal to which explains why we have the faculty of imagination (Millikan, 1984). Daydreaming and fantasy, along with imaginative involvement with fictions, are made possible by a system that already exists

for other purposes: the running of simulations" (Currie 1995b 158). Currie describes simulation as follows:

> ... [R]unning a test in the simulator should be a bit like having the experiences we would have if we really were acting out the strategy. Not quite like it, of course. That would be difficult, and dangerous too, because it would amount to the conditions of an hallucination. A compromise would be to reproduce the kinds of affective states we would have if the test were real. In the real case affective states are mediated by our beliefs and desires; we feel the sensations of fear not because the tiger is about to attack us but because we believe it is and desire it not to. So what the simulator does is to retain the connections between inner representation and bodily sensation that would be there if the representation was functioning as a belief; what is not retained is the belief-like connection to behaviour (1995b, 157).

Currie's hypothesis that strategy testing is the proper function of the imagination has been adopted by some researchers of quite different theoretical persuasion (e.g., Carruthers 1996, 266). On our view, however, this hypothesis is overly restrictive. Currie's proposal entails that the function of the pretense mechanisms is limited to considering possible courses of action or strategies to take. But one might also use the imagination and pretense mechanisms described in our theory to work on problems that are not strategies for action. For instance, one might use the mechanisms to figure out what will happen to one's domicile if there's a great deal of rainfall. Thus, on our view, it's plausible that the 'proper function' of the pretense mechanisms is more broadly for hypothetical and counterfactual reasoning.

5.2. Metarepresentational Accounts of Pretense

The second cluster of theories of pretense that we will consider are those in which a special kind of metarepresentation plays a central role. The most influential of these is the theory developed in a series of publications by Alan Leslie and his collaborators

(Leslie 1987, 1994; Leslie and Thaiss 1992; Leslie and Roth 1993). As we see it, Leslie's theory can be divided into two quite distinct parts. One of these parts is comfortably compatible with the theory we've proposed in Section 4 though the other is not.

A central concern of Leslie's theory is the avoidance of what he calls "representational abuse." Pretense emerges quite early in children, at a time when their knowledge of the world and their understanding of terms and concepts is still quite fragmentary. But when children pretend that a banana is a telephone or that an empty cup is full of tea, it is plausible to suppose they are relying on mental representations that "distort reality" -- they represent states of affairs that are simply not the case. "Why," Leslie asks, "does pretending not undermine their representational system and bring it crashing down?" (1987, p. 412) Why, for example, doesn't the child end up believing that some bananas are telephones or that cups can be full when they clearly look to be empty? The phenomenon that Leslie is concerned with here is, of course, just a special case of the one that we noted in Section 3.4. In both children and adults, episodes of pretense have only a quite limited impact on the post-pretense cognitive states of the pretender, and these effects are very different from what they would be if the pretender really believed the contents of the pretense. So a theory of pretense must explain how the representations that underlie episodes of pretense are "marked off, or 'quarantined,' from the primary representations" (1987, p. 415) whose job it is to provide a "literal and 'sober'" account of the world. (1987, p. 414) The infant must "have some way of marking information from pretend contexts to distinguish it from information from serious contexts."(1987, p. 416) As Leslie notes, one strategy for explaining how children manage to avoid representational abuse would be to posit a separate code or representational format for pretense-subserving representations. But this would make it significantly more difficult to explain the fact that pretenders can use their general knowledge and background beliefs to draw inferences within an episode of pretense since the pretense subserving representations and the background beliefs would be stored in two different codes. What Leslie proposes instead is that the representations that subserve pretense be "marked" in a special way to indicate that their functional role is

different from the functional role of unmarked (or "primary") representations. To use the terminology that Leslie employs, these marked representations are *decoupled* copies of the primary representations which do not have the "normal input-output relations" (417) that unmarked primary representations have.

The notational device that Leslie uses to mark pretense subserving representations is to enclose them in quotation marks, and since quotation marks are standardly used to form names of the representational expressions that they enclose, Leslie initially called these marked representations *metarepresentations*. This, however, proved to be an unfortunate choice of terminology which provoked a great deal of misunderstanding and criticism, much of it turning on the question of whether the young children, to whom Leslie attributed these marked representations, actually had the *concept* of representation, and therefore could think of a representation *as* a representation. If they couldn't, critics urged, then the marked representations were not really *meta*representations at all. (Perner 1988, 1991, 1993)[15] Leslie's response to these objections was to insist that he intended "metarepresentation" as a technical term for representations that played the role specified in his theory, and that the theory did not claim that people who had these representations conceived of them as representations of representations. To avoid further confusion, he abandoned the term 'metarepresentation' in favor of the more obviously technical term 'M-representation'.

Once these terminological confusions are laid to rest, it should be clear that the part of Leslie's theory that we have sketched so far is actually very similar to part of the theory that we have been defending. For, as we noted in 5.1.2, to claim that class of representations is specially marked and that the marking has important consequences for how the representations are treated is another way of saying that marked representations and unmarked representations are functionally different. Since the "box" metaphor is just a notational device for distinguishing representations that have systematically different

[15] For instance, Perner writes, "In our use of the term 'metarepresentation' we should adhere to Pylyshyn's (1978, p. 593) implicit definition as *representing the representational relation itself*, because his definition is unambiguous, and he was one of the first to use this term in cognitive science" (Perner 1993, 114-5).

functional or computational properties, Leslie's hypothesis that representational abuse is avoided because the representations subserving pretense are "quarantined" or "marked off" (1987, p. 415) is equivalent to claiming, as we do in our theory, that pretense-subserving representations are in a box of their own.[16] Another point of similarity between Leslie's theory and ours is that it does not posit a separate code or system of representation for the cognitive processes underlying pretense. The representations in the Possible World Box (in our theory), or within the quotation marks (in Leslie's theory) are tokens of the same types as the representations in the Belief Box (to use our preferred jargon) or in the pretender's primary representations (to use Leslie's). Also, in both theories the pretender's real beliefs (or "general knowledge" (1987, p. 419)) can be used as premises in elaborating an episode of pretense, and the inference mechanism that is responsible for these elaborations is the same one that is used in reasoning using only real beliefs. Leslie does not address the problem of avoiding contradictions between general knowledge and pretense assumptions, nor does he offer an account of the motivation for the behavior produced in pretense. So there are mechanisms and processes posited in our theory without any analogs in Leslie's account. Nonetheless, the part of Leslie's theory that we have set out so far can plausibly be viewed as simply a notational variant of part of our theory. Actually, though, the point should be made the other way round: Part of our theory is a notational variant of part of his -- and this is no accident since Leslie's work has been a major influence on our own.

A second central theme in Leslie's theory, and one that does not fit comfortably with ours, is the claim that "pretend play ...[is] a primitive manifestation of the ability to conceptualize mental states"(1987, p. 424) and thus that "pretense is an early manifestation of what has been called *theory of mind.*" (1987, p. 416) Because he thinks that pretense involves some understanding or conceptualizing of mental states, and also because he sees a close parallel between the "semantic properties of mental state

[16] Leslie actually disavows the Pretense Box hypothesis. The clearest place this comes out is, ironically, in a paper that Leslie co-authored with us (Nichols et al. 1996, 56). However, the existence of a Pretense Box is entirely compatible with the part of Leslie's theory that we've described thus far. Leslie's rejection of a Pretense Box depends, rather, on the second part of his theory, which we begin to discuss in the next paragraph.

expressions" (like 'believe', 'expect' and 'want') and the "basic form[s] of pretense" Leslie thinks that "mental state expressions can provide a model with which to characterize the representations underlying pretend play." (1987, p. 416) In developing this idea, Leslie proposes "a second extension to the primary code" to work in conjunction with the quotation marks whose job it is to quarantine the pretense-subserving representations. Here is how Leslie elaborates this idea:

> Language has its mental state terms that denote relationships between agents and opaque propositions. In fact, the verb *pretend* is just such a term. I can add to my model formal elements that play a corresponding role in underlying mental representation. The second extension to primary code will be an item, **PRETEND**, representing an informational relation. This relation will hold between whatever primary representations of agents (e.g. mother) the infant has and decoupled expressions. Pretend metarepresentations might thus have the general form: **Agent-Informational Relation-"expression." Agent** ranges over, for example, persons and self, whereas **"expression"** can be filled by any decoupled representation. (1987, p. 417)

As an illustration, Leslie proposes that one of the mental representations underlying the tea party pretense might have the form: **I PRETEND "this empty cup contains tea."** (1987, p. 420) On our view, this second component of Leslie's theory -- the hypothesis that pretense-subserving mental representations always include both a specification of an Agent and a representation of the "informational relation" in which the Agent stands to the decoupled representations that indicates what is going on in the pretense -- is unnecessary and unwarranted. The theory of pretense doesn't need it to explain the evidence and would be better off without it. In defending our view it will be useful to begin by comparing Leslie's two part theory with the theory that we set out in Section 4.

Let us start with some facts that are not in dispute. As we noted in Section 4, adults and older children typically have beliefs about what they are pretending (and about what others are pretending, when they are engaged in collaborative pretense), and they

can report these beliefs when asked, using sentences like "I am pretending that this empty cup contains tea". On our theory, there are two quite distinct mental representations implicated in these first person reports. First, there is a pretense-subserving representation in the PWB whose content is (roughly): *This* [empty] *cup contains tea.* Second, since adults and older children can monitor their own pretense and form true beliefs about what they are pretending, there is a representation in the Belief Box whose content is (roughly) *I am pretending that this* [empty] *cup contains tea.* Though the former representation is an important part of the causal process that leads to the formation of the latter, it is only this latter representation that is directly responsible for a subject's verbal report about what she is pretending. The former representation, by contrast, is the one that is directly implicated in the production of pretense behavior, the drawing of inferences about what is happening in the pretense, etc. Note that this pretense behavior can include *verbal behavior*, e.g., "Should I pour some tea?" In principle, of course, the pretense could proceed without the subject having any beliefs with contents like *I am pretending that this (empty) cup contains tea.* Indeed, on our theory, the pretense could proceed perfectly well even if the subject did not have the *concept* of pretense and thus could have no beliefs at all with contents of the form: *I am pretending that p.*

There is, on our theory, a close parallel between beliefs and reports about pretense, on the one hand, and beliefs and reports about desires, on the other.[17] Just as adults and older children have beliefs about what they are pretending and can report those beliefs, so too they typically have beliefs about their desires, particularly those desires that are currently guiding their behavior. On our theory, there are typically two quite distinct mental representations implicated in the causal process leading a subject to make a report like "I want to drink some water." First, there is the representation that subserves the desire itself. This representation, which is located in the subject's Desire Box has the content: *I drink some water.* Second, there is a representation in the subject's Belief Box whose content is *I want to drink some water.* As in the case of pretense, the first of these representations is an important part of the causal process that leads to the formation of the

[17] In a recent paper, Currie (forthcoming) has explored this parallel in some detail, and our development of the point has been significantly influenced by Currie's discussion.

second. But only the second representation, the one in the Belief Box, is directly involved in producing the subject's verbal report. By contrast, it is the representation in the Desire Box (in conjunction with various beliefs about the environment) that leads the subject to reach for the glass of water in front of her and raise it to her lips. And, just as in the case of pretense, the process that leads to drinking could proceed perfectly well even if the subject did not have the concept of *wanting* and thus could have no beliefs at all of the form: *I want that p*. So, on our theory, it is entirely possible that young children, or non-human primates, have lots of beliefs and desires though they have no theory of mind at all and are entirely incapable of conceptualizing mental states.

On Leslie's theory of pretense, the parallel that we have drawn between desiring and pretending breaks down. For Leslie, all episodes of pretense are subserved by representations of the form: **I PRETEND "p"**. Thus, while Leslie would agree that an agent can have desires and act on them without having the concept of desire, his theory entails that an agent *cannot* engage in pretense without having the concept of pretense. (He also seems to think that an agent cannot engage in pretense without believing that she is pretending.) As we see it, however, there is no more reason to suppose that young children who pretend have the concept of pretense (Leslie's **PRETEND**) than there is to suppose that young children who have desires have the concept of desire. We attribute this latter concept to older children and adults not because they *act* on their desires but rather because they *talk* about desires and indicate in various ways that they are reasoning about them. Since young children can pretend without talking about pretending or indicating that they are reasoning about pretending, the claim that they have the **PRETEND** concept seems unwarranted. (See Harris & Kavanaugh 1993, 75 for a similar argument.)

Why does Leslie think that pretense is "a primitive manifestation of the ability to conceptualize mental states"(1987, p. 424) and that a representation involving the **PRETEND** concept underlies all episodes of pretense? As best we can tell, he has three arguments for this view. One focuses on the fact that individual and group pretense emerge at the same time; a second focuses on the parallels between pretense and mental

state expressions; the third turns on the fact that autistic children are poor at mind reading tasks and do not exhibit normal pretend play. We are not convinced by any of these arguments, and in the remainder of this section we propose to say why. Before doing so, however, we need to take up another point which makes our critical project a bit more complicated. Though Leslie is more attentive to issues of cognitive architecture than many authors, there is one crucial point on which his theory is architecturally less explicit than one would like. He insists that episodes of pretense are subserved by representations of the form: **Agent PRETEND "expression"**, but he never explains where in the mind these representations are *located.* Since we are reasonably confident that Leslie would accept Figure 1 as a plausible sketch of the basic boxology of the mind, and since we know that he rejects the Pretense Box hypothesis[18] there appear to be two possibilities. Either the representations are located in the Belief Box or they are located in the Desire Box. In setting out our case against the second part of Leslie's theory we'll start by assuming that his theory would locate them in the Belief Box. Though most of what we say will not depend on this assumption, it will make the discussion more concrete and easier to follow. At the end of this section we'll note some of the special problems that arise if Leslie wants to claim that the **Agent PRETEND "expression"** representations are located in the Desire Box.

The first argument we want to consider is aimed quite explicitly at theories like ours on which pretending does not require the concept of pretense (just as desiring does not require the concept of desire). If this were true, Leslie maintains, it would be entirely possible for a child to engage in solitary pretense without being able to engage in pretense with another person or understanding what the other person was doing when she pretends; but, Leslie's argument continues, as a matter of empirical fact this simply does not happen (Personal communication & Leslie 1987, pp. 415-416). Children begin to pretend by themselves and to engage in joint pretense at exactly the same time. Theories like ours, Leslie argues, have no explanation for this important empirical fact, while his theory has an obvious explanation. If engaging in pretense and understanding pretense in

[18] See footnote 16.

others both depend on representations that include the **PRETEND** concept, then neither will be possible until that concept becomes available.

We have a pair of concerns with this argument, one of them is primarily conceptual, while the other is largely empirical. We'll start with the conceptual point. What is it to *understand* what another person is doing when she pretends that p? There are, it seems, two quite different accounts that might be offered. On the first, which we'll call the *behavioral account*, what one understands is that the other person is *behaving in a way that would be appropriate if p were the case.* On the second, which we'll call the *mentalistic account*, what one understands is that the other person is behaving in a way that would be appropriate if p were the case *because she is in a particular mental, viz. pretending that p.* Now, as Leslie notes, if a child has no understanding at all of pretense, then pretense behavior will often seem utterly bizarre and puzzling (1987, p. 416). (Why on earth would Moma be talking to a banana?!) But by the age of 2 or even earlier children obviously see nothing puzzling about pretense behavior. Quite the opposite; when Moma pretends that the banana is a telephone, they plunge right in and join the pretense. But, and this is the crucial point, in order to do this the child needs no more than a *behavioral* understanding of pretense. In order to engage in the banana / telephone pretense, the child must understand that Moma is behaving in a way that would be appropriate if the banana were a telephone. But the child need not have a *mentalistic* understanding of pretense. Indeed, a child with a behavioral understanding of pretense could engage in a quite elaborate two-person pretense *without understanding that the other person has any mental states at all.* So, from the fact that a child engages in two-person or group pretense it does not follow that the child is exhibiting "a primitive manifestation of the ability to conceptualize mental states." Participation in two-person pretense does not support Leslie's claim that the pretender has mental representations of the form: **Agent PRETEND "expression"**.

Let us now turn to the empirical issue. Leslie claims that an understanding of pretense in others emerges at the same time as the ability to engage in pretense oneself. Is this true? In light of the distinction between *behavioral* and *mentalistic* ways of

understanding pretense, it should be clear that the claim is ambiguous. It could mean that pretense behavior appears at the same time as a *behavioral* understanding of pretense, or that pretense behavior emerges at the same time as a *mentalistic* understanding. With the former claim, we have no quarrel. Though, as we've just argued, it lends no support to the second part of Leslie's theory. If, on the other hand, what Leslie claims is that pretense behavior and a mentalistic understanding of pretense emerge at the same time, then it is far from clear that this claim is supported by the facts. As we've seen, the mere fact that children engage in two-person pretense is not enough to establish that they have a mentalistic understanding. Moreover, several investigators have maintained that a mentalistic understanding of pretense emerges gradually and is not fully in place until some years after the child begins to engage in pretense behavior.[19] If this is correct, then the empirical claim in Leslie's first argument is not merely unsupported; it is false.

Leslie's second argument for the "second extension" part of his theory turns on what he takes to be the close parallel between the semantic properties of mental state expressions like *believe, expect* and *want* and the "basic forms" of pretense. The parallels that Leslie has in mind are the following: (a) Mental state expressions create referentially opaque contexts. Even if the sentences 'Mrs. Thatcher is the Prime Minister' and 'Sarah-Jane believes that the Prime Minister lives at No. 10 Downing Street' are both true, it does not follow that the sentence 'Sarah-Jane believes that Mrs. Thatcher lives at No. 10 Downing Street' is true. This, Leslie maintains, is parallel to the fact that "object substitution" takes place in pretense. When a child pretends that a stick is a horse, it does not follow that one can freely substitute 'stick' for 'horse' in attributing psychological states to the child. So, even though sticks are inanimate, a child can pretend that a stick is a horse without pretending or believing that a horse can be inanimate. (b) "Propositions involving mental state terms do not logically imply the truth (or falsehood) of propositions embedded in them." (1987, p. 416) Thus, for example, the sentence 'John believes that the cat is white' does not entail that 'The cat is white' is true (or that it is false). This, Leslie suggests, is parallel to the fact that a person can pretend that a cup is

[19] For instance, Lillard's (1993) results suggests that children as old as four years think that someone can pretend to be a rabbit without knowing anything about rabbits. (See

empty whether or not the cup actually is empty. (c) "Assertions involving mental state terms do not logically entail the existence of things mentioned in the embedded proposition." (1987, p. 416) So the sentence 'Jacqueline believes the king of France is bald' can be true even though the king of France does not exist. The parallel here, according to Leslie, is that a person can pretend that something exists when it does not.

Though there are various quibbles that one might raise with the parallels Leslie notes, we are prepared to grant that they are real enough. However, it is our contention that Leslie has failed to see the real explanation of these parallels and that the conclusion that he wants to draw from them is implausible. As we see it, the explanation for the parallels that Leslie notes is that pretending, believing, wanting and expecting are all *propositional attitudes*, and that 'pretend', 'believe', 'want' and 'expect' are all propositional attitude *terms*. All the facts about pretense mentioned in the previous paragraph have exact parallels for believing, wanting and expecting. One can, for example, want it to rain tomorrow (or believe that it will) whether or not it rains tomorrow. And one can want to meet Santa Claus (or expect to) whether or not Santa Claus exists. Similarly, all the facts that Leslie notes about mental state terms like 'want', and 'believe' have exact parallels for 'pretend'. So the deep parallels are not those between pretending and the *terms* for propositional attitudes, but between pretending and propositional attitudes themselves, and between the term 'pretend' and other propositional attitude terms. Once this is seen, it makes Leslie's proposal to add '**PRETEND**' to the mental representations subserving pretense look very odd indeed. For if it is plausible to suppose that the mental representation subserving the pretense that a certain cup contains tea has the form: **I PRETEND "this empty cup contains tea,"** then, given the parallels we have noted, the mental representation subserving the belief that this cup contains tea should be: **I BELIEVE this cup contains tea**, and the mental representation subserving the desire that it rain tomorrow should be: **I DESIRE that it rain tomorrow**. And if this were the case, then it would impossible to believe *anything* without having the concept of belief, and impossible to desire *anything* without having the concept of desire. So any organism that had any beliefs and desires at all would have

---

also Jarrold, et al. 1994; Lillard & Flavell 1992; Lillard 1996).

to have these concepts and thus at least the beginnings of a theory of mind. The way to avoid this package of unwelcome conclusions is clear enough. We should buy into the first half of Leslie's theory (which, it will be recalled, is a notational variant on part of our theory) and reject the second half.

There is one further argument that figures prominently in Leslie's defense of his theory of pretense. The argument turns on the interesting and important fact, discovered by Leslie and others (Baron-Cohen et al. 1985; Leslie & Roth 1993) that autistic children typically exhibit a pair of quite striking deficits: The first is that their ability to engage in pretend play is severely impaired when compared with other children of the same mental age. The second is that their performance on false-belief understanding tasks and other standard tasks that are taken to indicate an understanding of mental states is also severely impaired when compared with other children of the same mental age. This suggests that the processes underlying pretense and the processes underlying our ability to understand mental states share some common mechanism. Leslie's hypothesis is that the impairment is in the decoupling mechanism. This, it will be recalled, is the mechanism that marks mental representations with quotation marks to indicate that they do not have the same functional role that these representations would have if they were subserving belief. In our version of the theory, what Leslie calls "decoupling" is accomplished by putting the representations in the Possible World Box. In order for it to be the case that a defect in the decoupling mechanism (or the system that puts representations into the PWB) leads to an impairment in theory of mind skills, it must be the case that decoupling (or putting representations in the PWB) plays a central role in understanding and reasoning about mental states. This is an intriguing and important hypothesis which Leslie develops and which we have discussed elsewhere (Nichols & Stich forthcoming). What is important, for present purposes, is that if the hypothesis is right (and we think it is) it offers no support at all for what we have been calling the second half of Leslie's theory of pretense. If the decoupler (or the system that puts representations into the PWB) is impaired then we would expect to find deficits in the ability to pretend, no matter what account one favors about the exact form of the representations that subserve pretense. And if the decoupler (or the system that puts representations into the PWB) also plays a

central role in reasoning about the mind and solving theory of mind tasks, then we should also expect deficits in this area no matter what account one proposes about the exact form of the representations subserving reasoning about the mind. So the facts about autism are simply irrelevant to the second half of Leslie's theory, which claims that the representations subserving pretense have the form: **I PRETEND "p"**.

Thus far we have been assuming that the **I PRETEND "p"** representation that Leslie takes to be essential to pretense are stored in the pretender's Belief Box. There is, however, an obvious problem with this assumption that is quite independent of the other problems we have raised for Leslie's theory. For if the **I PRETEND "p"** representation is in the Belief Box, then the theory gives us no explanation at all about why pretenders *do* anything. Representations in the Belief Box do not motivate behavior. That is the job of desires, not beliefs. So let us now explore the possibility that pretense might be subserved by **I PRETEND "p"** representations in the Desire Box. If this is the case then when someone pretends that a banana is a telephone he has a representation with the content *I pretend that the banana is a telephone* in his Desire Box. But this leads to a problem that is a mirror image of the problem we just raised for the idea that pretense subserving representations are in the Belief Box. Having the representation in the Desire Box might explain why the pretender does something, but on Leslie's view (though not on ours) agents always know that they are pretending, and this knowledge requires a suitable representation in the Belief Box. How does the representation get there? It might be thought that self-awareness of what is going on in the Desire Box would be enough to answer this question. But that would be a mistake. For the mechanism subserving self-awareness will produce a representation with the content *I want to pretend that the banana is a telephone* in the Belief Box, while what is needed is a representation with the content *I pretend (or I am pretending) that the banana is a telephone.*

Yet another problem with the view that Leslie's pretense subserving **I PRETEND "p"** representations are located in the Desire Box is that it does not explain how the pretender knows what to do when he wants to pretend that p. The mere desire to pretend

that p is not enough. It must be combined with appropriate beliefs in the Belief Box, of the form:

(1) *In order to pretend that p, one must (or can) do actions a, b, & c.*

That belief, along with the desire to pretend that p, will produce the desire to a & b & c. But where do beliefs like (1) come from? The most plausible answer is that the pretender has *a theory of pretense* that supplies them. Presumably, the theory says that to pretend that p you have to behave in a way that is similar to the way you would behave if p were the case. But if this is the 'analysis' of pretense that the theory provides, then this part of Leslie's theory again begins to look like a notational variant of ours. For the desire with the content *I pretend that p* is equivalent to the desire with the content *I behave in a way that is similar to the way I would behave if p were the case*, and that, near enough, is the same desire that motivates pretense behavior on our account. But, of course, once the **I PRETEND "p"** in the Desire Box has been analyzed in this way, the account of the mental representations subserving pretense lends no support at all to Leslie's contention that pretense is an early manifestation of the ability to conceptualize mental states.

Where does all of this leave us? As we see it, the second part of Leslie's theory – the part that maintains that all episodes of pretense are subserved by representations of the form: **I PRETEND "p"** — is beset by difficulties on every side. The empirical evidence that would be needed to support the claim is not there; the analogy between pretense and propositional attitude verbs is not a good one; the argument from autism is of no help; and when we start to think about the issue of *where* in the mind these representations are supposed to be located, all the options lead to problems. All of these difficulties can be avoided if we drop the second part of Leslie's theory and stick with the first part. And that part, it will be recalled, is fully compatible with the theory we developed in Section 4.

5.3. Perner's Account

We began this section by claiming that most recent accounts of pretend play are either "on-line simulation" accounts or "metarepresentational" accounts. However, Josef Perner's account may not fit either of these models. Perner's account of pretend play, like Leslie's account, is motivated by the fact that in order to avoid confusion, pretend-representations need to be "quarantined" from representations of reality. According to Perner, this problem can be solved without claiming that young children have any understanding of mental states. Rather, Perner suggests, "Pretend representations are not representations of the world as it is but of the world as it might be" (1991, 59). Perner elaborates this suggestion by exploiting an analogy with temporal contexts. He writes:

> The same need for quarantining exists when information about different times is to be stored. It won't do to mix "I am 2 years old" with "I am 3 years old." To avoid confusion about one's age, one must mark one of these representations as "past" or the other as "future." Analogously, my suggestion for pretense is to mark off the pretend scenario as "nonreal" or "hypothetical." So, although the need for quarantine is clear, it is not clear why quarantining requires metarepresentation. The need for quarantine is served adequately by multiple models representing different situations (1991, 60-61; see also 1991, 54-56; 1993, 117-118).

Perner's proposal can be interpreted in a number of ways. One possibility is that Perner thinks that pretend representations are distinguished from beliefs in virtue of differences in *content*. That is, perhaps Perner is suggesting that pretend representations and beliefs are simply *about* different things. If this is Perner's view, then it strikes us as distinctly unpromising, since a belief and a pretense might well represent the same situation. For instance, Bill might believe that the earth is flat, while Sue pretends that the earth is flat. In this case, the mental states seem to represent the same situation, although one representation is a pretense and the other is a belief. Indeed, as we've seen, it even seems possible for the same individual to pretend that the cup is empty and at the

same time believe that the cup is empty. In such cases, it's especially difficult to see how one can distinguish belief representations from pretense representations in terms of what the representation is about.

Another possible interpretation of Perner's proposal is that beliefs are distinguished from pretense representations in terms of *functional* differences rather than differences in content. That is, perhaps Perner is suggesting that pretenses and beliefs have different functional roles. If this is Perner's view, then his theory can plausibly be viewed as a notational variant of our theory. We have suggested that the crucial distinction between pretense representations and belief representations turns on the functional roles of the mental states, not on the contents of the representations. This functional distinction is captured 'boxologically' by positing a Possible World Box, with a distinctive set of connections to other mechanisms, in which the representations subserving pretense are elaborated. Perhaps Perner has this view in mind.

A third possible interpretation is that Perner is suggesting that pretend representations are distinguished from beliefs in terms of *both* function and content. If this is the right interpretation, then Perner's view actually resembles Leslie's view in important respects. As we argued in section 5.2, on Leslie's account, pretense representations differ in both content and function from beliefs. Perhaps Perner is proposing something similar when he suggests that the way to quarantine pretense from belief is to "mark off the pretend scenario as 'nonreal' or 'hypothetical'" (1991, 61). However, if this is Perner's view, the view seems problematic. As noted above, there is no reason to think that a pretend representation can't have the same content as a belief. Furthermore, adopting a functional distinction between pretenses and beliefs suffices to quarantine pretenses from beliefs. Positing an additional difference at the level of content only raises further problems.

All three of these interpretations of Perner seem possible. As a result, we're not quite sure how to graft his view on to the architecturally more explicit framework in which our theory is elaborated. From our perspective, the most promising reading is that

Perner is trying to distinguish belief from pretense in terms of the functional roles of these mental states. If so, then he is trying to capture the very distinction that we draw by positing a separate Possible World Box.

6. Conclusion

Despite the length of this paper, we think that we have only provided a bare sketch of a theory of pretense. Nonetheless, our account is considerably more explicit than the other accounts in the literature. By way of conclusion, we would like to recapitulate the main features of our theory and make some suggestions for further research. At the core of our theory is the idea that pretense representations are contained in a separate workspace, a Possible World Box which is part of the basic architecture of the human mind. In other words, pretense representations play a distinctive functional role in the cognitive economy. Pretense representations are not distinguished from beliefs in terms of the content of the representations. Indeed, in an important sense, pretense representations are in the same representational "code" as beliefs – in pretending, the set of representations being built up in the PWB is inferentially elaborated and updated by the same inference and UpDating mechanisms that operate over real beliefs. In addition to inferential elaboration, pretenders also elaborate the pretense in non-inferential ways, exploiting what we have called the Script Elaborator. All of this cognitive architecture is, we think, essential to imagination and pretense, but it does not explain why children actually enact the pretend scenarios. We suggest that the motivation for pretend play derives not from a "pretend desire", but from a real desire to act in a way that fits the description being constructed in the PWB. Finally, while our account does not claim that pretense requires mind-reading or theory of mind capacities, the account does leave open the possibility that pretense and mind-reading capacities use some of the same mechanisms, and thus that breakdowns would be correlated.

This possible connection between mind reading and pretense is one of the many issues that require further research. We think it's quite possible that the PWB is

implicated in mind reading skills, and one way to investigate this is to explore whether there are correlations between pretense and attribution. For example, if Mary is asked to pretend that trees are conscious, does the description built up in the PWB coincide with the attributions she would make if asked to characterize the beliefs of someone who thinks that trees are conscious? In addition to mind-reading skills, pretense mechanisms may be implicated in a number of other mental processes. We have suggested that counterfactual reasoning depends on the PWB and the UpDater, and it is important to investigate whether deficits in the capacity for pretense are correlated with deficits in the capacity for counterfactual reasoning. Pretense mechanisms may also be implicated in conditional planning, moral psychology, and aesthetic judgment, and we hope to explore some of these connections in future work. It will also be important to explore ways in which the mechanisms of pretense are distinct from other mental mechanisms. For instance, we need to be able to distinguish pretense from a variety of other sorts of processes that have been discussed in the literature, e.g., self-deception. Further, our model suggests that we need to explore whether there are systematic differences between the way that the UpDating mechanisms process the contents of the PWB and the contents of the belief box; similarly we need to explore whether there are systematic differences in how the emotional response systems process the contents of the PWB and the contents of the belief box. Finally, there are a host of issues about the individual's understanding and awareness of pretense itself. What mechanisms underlie the awareness of the contents of the PWB? And what exactly is involved in conceptualizing pretense? We suspect that the capacity to be aware of the contents of the PWB is independent of the capacity to understand pretending. But there is a considerable amount of conceptual and empirical work to be done here. We have only offered a few suggestions of some issues that merit further attention. The crucial point is that developing an architecturally explicit theory of pretense should help determine which empirical and conceptual issues need to be explored to enrich our understanding of pretense and imagination.

Acknowledgments:

We would like to thank the participants in our study. We would also like to thank Gregory Currie, Alison Gopnik, Paul Harris, Barbara Landau, Alan Leslie, Angeline Lillard, Josef Perner, Richard Samuels, Brian Scholl and Jonathan Weinberg for comments on an earlier version of this paper.

References

Abelson, R. 1981. Psychological status of the script concept. *American Psychologist,* 36, 715-729.

Baron-Cohen, S. 1995. *Mindblindness: An essay on autism and theory of mind.* MIT Press.

Baron-Cohen, S., Leslie, A. and Frith, U. 1985. Does the autistic child have a "theory of mind"? *Cognition*, 21, 37-46.

Bartsch, K. & Wellman, H. 1995. *Children talk about the mind.* Oxford University Press.

Buss, D., Larsen, R., Westen, D., & Semmelroth, J. 1992. Sex differences in jealousy: Evolution, physiology, and psychology. *Psychological Science*, 3, 251-255.

Carruthers, P. 1996. "Autism as mind-blindness: An elaboration and partial defence". In P. Carruthers & P. Smith (eds.), *Theories of Theories of Mind.* Cambridge University Press.

Carruthers, P. and Smith, P. 1996. *Theories of theories of mind.* Cambridge: Cambridge University Press.

Currie, G. 1990. *The nature of fiction.* Cambridge University Press.

Currie, G. 1995a. *Image and mind.* Cambridge University Press.

Currie, G. 1995b. Imagination and simulation: Aesthetics meets cognitive science. In A. Stone and M. Davies (eds.) *Mental simulation: Evaluations and applications*, Oxford: Basil Blackwell.

Currie, G. 1995c. The moral psychology of fiction. *Australasian Journal of Philosophy*, Vol. 73, 250-259.

Currie, G. 1995d. Visual imagery as the simulation of vision. *Mind and Language* 10, 25-44.

Currie, G. 1996. Simulation-theory, theory-theory, and the evidence from autism. In P. Carruthers & P. Smith (eds.), *Theories of Theories of Mind.* Cambridge University Press.

Currie, G. 1997. The paradox of caring. In M Hjort and S. Laver (eds.) *Emotion and the arts.* Oxford University Press.

Currie, G. and Ravenscroft, I. 1997. Mental simulation and motor imagery. *Philosophy of Science*, Vol. 64, 161-180.

Currie, G. forthcoming. Pretense, pretending and metarepresenting.

Davies, M. and Stone, T. 1995a. *Folk psychology: The theory of mind debate.* Oxford: Blackwell.

Davies, M. and Stone, T. 1995b. *Mental simulation: Evaluations and applications.* Oxford: Blackwell.

Fein, G. & Kinney, P. 1994. He's a nice alligator: Observations on the affective organization of pretense. In A. Slade & D. Wolf (eds.), *Children at play: Clinical and developmental approaches to meaning and representation*, New York: Oxford University Press.

First, E. 1994. The leaving game, or I'll play you and you play me: the emergence of dramatic role play in 2-year-olds. In A. Slade & D. Wolf (eds.), *Children at play: Clinical and developmental approaches to meaning and representation*, New York: Oxford University Press.

Garvey, C. 1979. An approach to the study of children's role play. *The Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, 12.

Goldman, A. 1992. Empathy, mind, and morals. *Proceedings and Addresses of the American Philosophical Association, 66*, No. 3, pp. 17-41.

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

Goldman, A. 1993. The Psychology of folk psychology. *Behavioral and Brain Sciences*, 16, 15-28.

Gopnik, A. & Slaughter, V. 1991. Young children's understanding of changes in their mental states. *Child development*, 62, 98-110.

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

Gordon, R. 1986. Folk psychology as simulation. *Mind and Language*, 1, 158-170.

Gordon, R. & Barker, J. 1994. Autism and the 'theory of mind' debate. In G. Graham & G. L. Stephens (eds.), *Philosophical psychopathology: A book of readings*, Cambridge, MA: MIT Press.

Gould, R. 1972. *Child studies through fantasy.* New York: Quadrangle Books.

Harris, P. 1991. The work of the imagination. In A. Whiten (ed.), *Natural theories of mind.* Oxford: Blackwell.

Harris, P. 1994. Understanding pretense. In Lewis and Mitchell (1994). 235-259.

Harris, P. L. & Kavanaugh, R. D. 1993. Young children's understanding of pretense. *Monographs of the Society for Research in Child Development*, 58, No. 1.

Harris, P. 1993. Pretending and planning. In S. Baron-Cohen, H. Tager-Flusberg and D. Cohen (eds.), *Understanding other minds: Perspectives from autism.* Oxford University Press.

Harris, P. 1994. Thinking by children and scientists: False analogies and neglected similarities. In L. Hirschfeld and S. Gelman, *Mapping the mind: Domain specificity in cognition and culture.* New York: Cambridge University Press.

Harris, P. 1995. Imagining and pretending. In M. Davies and T. Stone (eds.), *Mental simulation.* Blackwell.

Izard, C. 1991. *Psychology of emotion.* New York: Plenum Press.

Jarrold, C., Carruthers, P., Smith, P., and Boucher, J. 1994. Pretend play: Is it metarepresentational?. *Mind & Language*, 9, 445-468.

Kunda, Z. 1987. Motivated inference. *Journal of Personality and Social Psychology*, 53, 636-647.

Leslie, A. 1987. Pretense and representation: The origins of "theory of mind." *Psychological Review*, 94, 412-426.

Leslie, A. 1991. The theory of mind impairment in autism: Evidence for a modular mechanism of development? In A. Whiten (ed.), *Natural theories of mind.* Oxford: Blackwell.

Leslie, A. 1994. Pretending and believing: Issues in the theory of ToMM. *Cognition*, 50, 211-238.

Leslie, A. and Roth, D. 1993. What autism teaches us about metarepresentation. In S. Baron Cohen, H. Tager-Flusberg, and D. Cohen (Eds.), *Understanding other minds: Perspectives from autism.* Oxford: Oxford University Press.

Leslie, A. and Thaiss, L. 1992. Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition* 43, 225-251.

Leslie, A., Xu, F., Tremoulet, P., and Scholl, B. 1998. Indexing and the object concept: developing 'what' and 'where' systems. *Trends in Cognitive Sciences* 2, 10-18.

Lewis, C and Mitchell, P. 1994. *Children's early understanding of mind: Origins and Development.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Lewis, D. 1986. *On the plurality of worlds.* NY: Basil Blackwell.

Lillard, A. 1993. Young children's conceptualization of pretense: Action or mental representational state? *Child Development*, 64, 372-386.

Lillard, A. 1994. Making sense of pretense. In Lewis and Mitchell (1994). 211-234.

Lillard, A. 1996. Body or mind: Children's categorizing of pretense. *Child Development*, 67, 1717-1734.

Lillard, A. & Flavell, J. 1992. Young children's understanding of different mental states. *Developmental Psychology*, 28, 626-634.

MacWhinney, B. & Snow, C. 1990. The child language data exchange system: An update. *Journal of Child Language*, 17, 457-472.

Millikan, R. 1984. *Language, thought, and other biological categories.* Cambridge, MA: MIT Press.

Mosher, D., Barton-Henry, M., and Green, S. 1989. Subjective sexual arousal and involvement: development of multiple indicators. *Journal of Sex Research*, 25, 412-425.

Nichols, S. and Stich, S. forthcoming. *Mindreading.* Oxford: Oxford University Press.

Nichols, S., Stich, S., and Leslie, A. 1995. Choice effects and the ineffectiveness of simulation: Response to Kuhberger et al.. *Mind & Language,* v. 10, no. 4, 437-445.

Nichols, S., Stich, S., Leslie, A., and Klein, D. 1996. Varieties of off-line simulation. In P. Carruthers and P. Smith (eds.), *Theories of theories of mind*, Cambridge: Cambridge University Press, 39-74.

Peacocke, C. 1983. *Sense and Content.* Oxford: Clarendon Press.

Perner, J. 1988. Developing semantics for theories of mind: From propositional attitudes to mental representation. In J. Astington, P. Harris, and D. Olson (eds.) *Developing theories of mind.* Cambridge University Press, 141-172.

Perner, J. 1991. *Understanding The representational mind.* MIT Press.

Perner, J. 1993. Rethinking the metarepresentation theory. In S. Baron-Cohen, H. Tager-Flusberg and D. Cohen (eds.), *Understanding other minds: Perspectives from autism.* Oxford University Press.

Perner, J., Barker, S. and Hutton, D. 1994. *Prelief:* The conceptual origins of belief and pretense. In Lewis and Mitchell (1994). 261-286.

Perry, J. 1993. *The problem of the essential indexical and other essays.* New York: Oxford University Press.

Piaget, J. 1962. *Play, dreams, and imitation in childhood.* Translated by C. Gattegno and F. M. Hodgson. New York: Norton.

Pylyshyn, Z. 1978. When is attribution of beliefs justified? *The Behavioral and Brain Sciences*, 1, 592-3.

Pylyshyn, Z. 1987. *The robot's dilemma: The frame problem in artificial intelligence.* Norwood, N.J.: Ablex.

Schank, R. and Abelson, R. 1977. *Scripts, plans, goals, and understanding : An inquiry into human knowledge structures* . Hillsdale, N.J. : L. Erlbaum Associates.

Smith, D. and Over, R. 1987. Correlates of fantasy-induced and film-induced male sexual arousal. *Archives of Sexual Behavior*, 16, 395-409.

Smith, D. and Over, R. 1990. Enhancement of fantasy-induced sexual arousal in men through training in sexual imagery. *Archives of Sexual Behavior*, 19, 477-490.

Sorensen, R. 1998. Self-strengthening empathy. *Philosophy and Phenomenological Research*, v. 58, no. 1.

Sterelny, K. 1990. *The representational theory of the mind: An Introduction.* Oxford: Blackwell.

Stich, S. and Nichols, S. 1992. Folk psychology: Simulation or tacit theory. *Mind & Language*, v. 7, no. 1, 35-71.

Stich, S. and Nichols, S. 1995. Second thoughts on simulation. In A. Stone and M. Davies (eds.) *Mental simulation: Evaluations and applications*, Oxford: Basil Blackwell, 87-108.

Stich, S. and Nichols, S. 1997. Cognitive penetrability, rationality, and restricted simulation. *Mind & Language*, 12, 297-326.

Vygotsky, L. 1967. Play and its role in the mental development of the child. *Soviet Psychology*, 5, 6-18.

Walton, K. 1990. *Mimesis as make-believe: On the foundations of the representational arts.* Harvard University Press.

Walton, K. 1997. Spelunking, simulation and slime: On being moved by fiction. In M Hjort and S. Laver (eds.) *Emotion and the arts.* Oxford University Press.

Wimmer, H. and Perner, J. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-128.