ELSEVIER

Cognition 74 (2000) 115–147

COGNITION

www.elsevier.com/locate/cognit

A cognitive theory of pretense

Shaun Nichols[a,*], Stephen Stich[b]

[a]*Department of Philosophy, College of Charleston, Charleston, SC 29424, USA*
[b]*Department of Philosophy, Center for Cognitive Science, Rutgers University, New Brunswick, NJ 08901, USA*

Received 2 November 1998; received in revised form 1 June 1999; accepted 5 October 1999

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 describing 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 mental 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.

Keywords: Pretense; Imagination; Theory of mind; Metarepresentation; Simulation; Cognitive architecture; Possible World

* Corresponding author. Fax: + 1-732-4456715.
E-mail address: nichols@cofc.edu (S. Nichols)

0010-0277/00/$ - see front matter
PII: S0010-0277(99)00070-0

1. Introduction

Pretend play in young children is so familiar and so natural that it is 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), and there is a large empirical literature on the issue (see Fein, 1981 for a review). 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. Currie, 1995b, 1997; Gordon & Barker, 1994; Harris, 1991, 1994b, 1995; Harris & Kavanaugh, 1993; Leslie, 1987, 1994; Lillard, 1993, 1994; Perner, Barker & Hutton, 1994).

Much of the recent debate over pretense has been fueled by work on the capacity for understanding other minds or ''mindreading''. Although most researchers agree that there is an important connection between pretend play and mindreading, there are quite fundamental disagreements about the mental mechanisms that make mindreading 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.[1] Writers on both sides maintain that their theory provides a better explanation of mindreading phenomena and of the deficits seen in autism (e.g. Carruthers, 1996; Currie, 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 mindreading 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 mindreading. 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 (Currie, 1995b; Goldman, 1992; Sorensen, 1998), moral understanding (Currie, 1995c), literary appreciation (Currie, 1990, 1995a,b; Walton, 1990, 1997), and visual imagery (Currie, 1995d). Although we have been frequent critics of off-line-simulation theory (Nichols, Stich & Leslie, 1995; Nichols, Stich, Leslie & Klein, 1996; Stich & Nichols, 1992, 1995, 1997), we think it is 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 is essential to have a plausible theory of the mental mechanisms that underlie the capacity for pretense itself.

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

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. It is our contention that all the other theories of pretense that have been proposed in the recent literature are underdescribed in important ways, and in particular that all of them tell us far too little about the sort of mental architecture 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. Since our theory will be much more explicit than previous accounts, it is also more likely to be mistaken. But that doesn't really worry us, since it is 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 experimental 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 briefly describe a few examples of pretense in children and adults, and draw attention to some of the features of these examples, 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 3, 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 2. Finally, in Section 4, 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 and features of pretense

Much of the literature on pretense is guided by two examples from the work of Alan Leslie. In one of these, which Leslie (pers. commun.) tells us he observed in one of his own children, a child uses a banana as if it were a telephone (Leslie, 1987). 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.'' The second example comes from a series of experiments in which 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, p. 223). When asked to point at the 'empty cup', 2-year-olds pointed to the cup that had been turned upside down (Leslie, 1994). The final example of childhood pretense that we'll mention

comes from a set of protocols of spontaneous pretense in children that was assembled by Gould (1972). Gould reports that one 3-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).

Since we want to develop a theory of pretense that can accommodate adult pretense as well as pretense in children, there is no need to restrict ourselves to examples of pretend episodes in children. Pretend episodes in adults will work just as well. As it happens, however, there are surprisingly few examples of adult pretense described in the psychological literature. So we set out to collect our own. To do this we had college student volunteers carry out a number of individual and group pretense scenarios. The descriptions of individual pretense scenarios that we asked our subjects to act out included:

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.

For two person 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.

After subjects carried out the scenarios, we had informal oral interviews with them. Where appropriate in the sections to follow, we'll present examples drawn from the transcripts of these episodes of pretense.[2]

Let us turn now to some of the features that can be found in the examples of pretense we've assembled. Since the features we will note 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 4, it can handle more of them than any of the other theories that have been proposed.

2.1. Getting pretense started: the initial premise

Typical episodes of pretense begin with an initial premise or set of premises, which are the basic assumptions 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 the example from Gould, the assumption is that the boy is a pussycat.

[2] For a detailed description of the method used in this study along with extensive quotes from the transcripts, see Nichols and Stich (in press b).

To get the pretense going the pretender must either produce the initial premise (if she initiates the pretense) or she must figure out what the initial premise is and decide whether or not she is going to proceed with the pretense (if someone else initiates 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.

2.2. 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.

2.3. 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 one of our fancy restaurant scenarios, the ''waiter'' provided the ''diner'' with a menu (actually a blank sheet of paper) and the diner then announced that he would have the chicken pasta for dinner. Our post-pretense interviews indicated that some of these elaborations cohere with the decisions and choices that the person pretending would actually make. In other cases, however, the elaborations depart from what the person would actually decide, sometimes quite radically. Thus, for example, in one of our fancy restaurant pretenses, the waiter pretended to decapitate one of the diners! A theory of pretense needs to be able to accommodate these kinds of elaborations as well as the more sober inferential elaborations.

2.4. Production of appropriate pretend behavior

Perhaps the most obvious fact about pretense is that pretenders actually *do* things, i.e. 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. 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. Our adult subjects did much the same, though they were 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?

2.5. 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. One 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 (1987) emphasizes, even very young children do not come to believe that bananas sometimes *really are* telephones. Nor, of course, do adults. 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 dead cats. 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. 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.

3. A theory about the cognitive mechanisms underlying pretense

3.1. 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) 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. Fig. 1 is a sketch 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. 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.

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.[3]

[3] 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.

3.2. 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 cognitive architecture – three new ''boxes'' that we propose to add to the account depicted in Fig. 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 possible world.[4] We are inclined to think that the mind uses the PWB for a variety of tasks including mindreading, strategy testing, and empathy. Although we think that the PWB is implicated in all these capacities, we suspect that the original evolutionary function of the PWB was rather to facilitate reasoning about hypothetical situations (see Currie, 1995b for a contrasting view). 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 Leslie's tea party pretense, the episode begins when a representation with the content *We are going to have a tea party* is placed in the PWB. 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. Thus, 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 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

[4] 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*. The issue becomes more complicated for *logically* impossible worlds that invoke obvious contradictions; we discuss this more fully in Nichols and Stich (in press b).

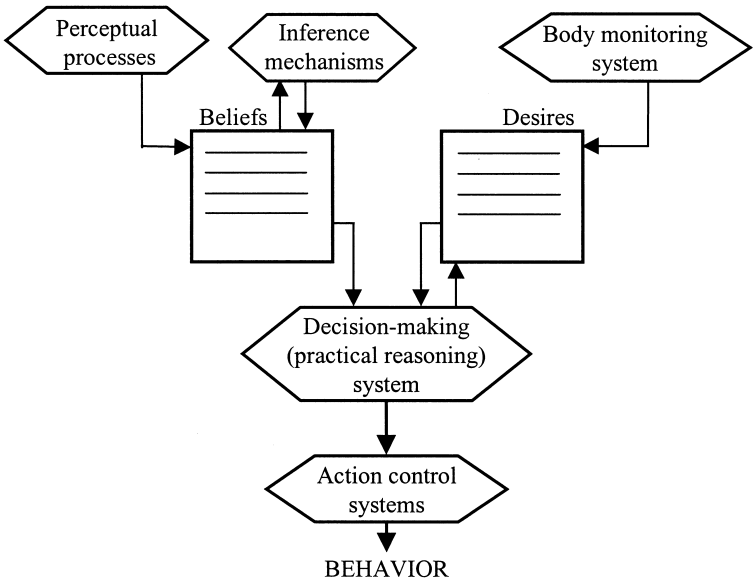

Fig. 1. Basic cognitive architecture.

PWB is the pretense initiating representation. From *We are going to have a tea party* there are relatively few 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 is going to generate glaring contradictions within the possible world description that is being built up in the Possible World Box. This would produce inferential chaos

in the Possible World Box, and obviously this does not happen. 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 is a perfectly ordinary event which requires that his cognitive system update a few of his beliefs. Other cases are more dramatic and complicated. 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, for example, 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 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 Fig. 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 sub-component 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*

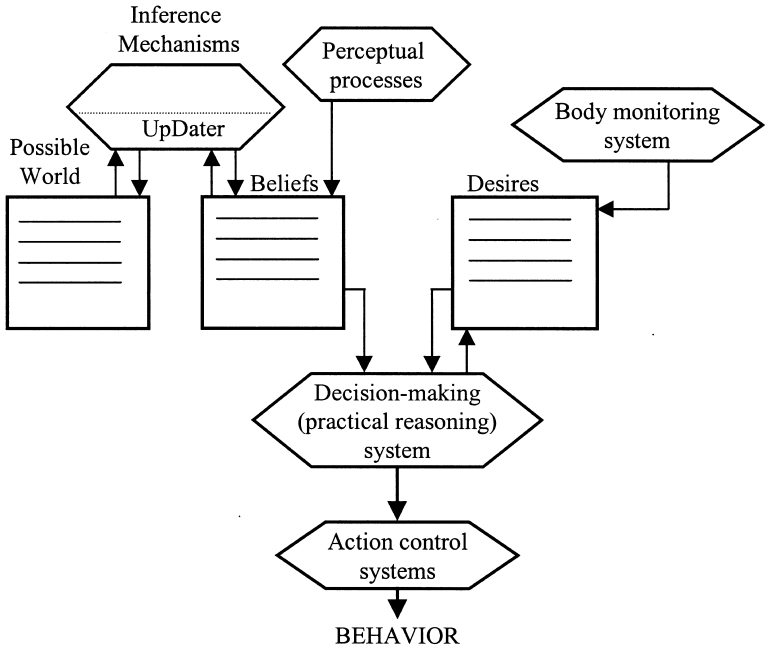

Fig. 2. An initial proposal for the cognitive architecture underlying pretense.

has been filtered out (i.e. altered or eliminated) by the UpDater. 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 properties and the logical form of pretense representations. One possibility that we find attractive is that the representations in the PWB have the *same logical form* as representations in the Belief Box, and that their representational properties are *determined in the same way*. When both of these are the case, we will say that the representations are *in the same code*.[5] Since mental processing mechanisms like the inference mechanism are

[5] There is much controversy about how the semantic properties of mental representations are determined (see, for example, Stich & Warfield, 1994). Our theory takes no stand on this issue apart from assuming that the same account will apply to both Belief Box representations and PWB representations. For more on the logical form of mental representations see, for example, Braine (1994) and Higginbotham (1995).

usually assumed to be sensitive to the logical form of representations, the inference mechanism will handle pretense representations and belief representations in much 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 much the same way that they process beliefs.

If we are right that pretense representations are in the same code as beliefs, then the pretense representation *Hamlet is the Prince of Denmark* will have 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. 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 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 substantial literature about indexicals and demonstrative thoughts in philosophy and cognitive science (e.g. Leslie, Xu, Tremoulet & Scholl, 1998; Peacocke, 1983; Perry, 1993). Perhaps, then, the UpDater and other components of the inference mechanism treat ''demonstrative pretenses'' in much the same way they treat ''demonstrative thoughts''.[6]

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 (for example see Abelson, 1981; Schank & Abelson, 1977). 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. So, for example, all the subjects who participated in our fast food restaurant scenario

[6] 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 be 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 architectural explicitness that it brings empirical issues like this into much sharper focus.

followed the standard pattern of the fast food restaurant script, i.e. order first, then pay and get the food, then eat.[7] 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. This still leaves many options open, however. In the protocols that we collected, sometimes the pretender's choices followed what the pretender would normally do in real life. But on other occasions, the pretender deviated from what he would normally do. 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?

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 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 component to our account of mental architecture, 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.

3.3. Explaining pretense behavior: information and motivation

Fig. 3 is a sketch of the cognitive mechanisms that we now have in place in our theory. Those mechanisms provide at least the beginnings of an explanation for several of the features of pretense set out in Section 3. 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.

[7] The script constraints are only ''soft'' constraints, however, and an imaginative pretender might elect to violate the script constraints quite dramatically. For example, in one of our fancy restaurant scenarios the ''waiter'' crushed 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. Also, sometimes the scripts are themselves not accurate descriptions of the world but, rather, stylized depictions. For example, the child's script for behaving like a train is to make the sound ''Chugga chugga, choo choo'', though he has in fact never heard a train make that sound. In a number of recent papers Paul Harris (1993, 1994a) has emphasized the importance of scripts and paradigms in pretense and imagination. We are indebted to Harris and to an anonymous referee for prompting us to think more about these matters (see also Bretherton, 1989).

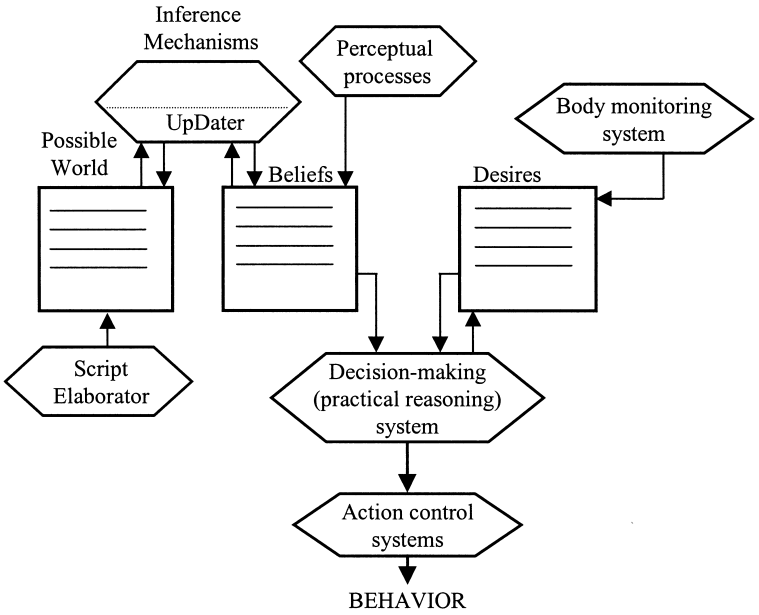

Fig. 3. Our theory of pertense.

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, 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, p. 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 $q_1\&q_2\&\ldots\&q_n$, where p is the pretense premise and $q_1\&q_2\&\ldots\&q_n$ 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''.[8]

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 pretender's 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 9-1-1. However, she took precautions to ensure 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.

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.

4. 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 2. A second central theme is that, when one sets out to elaborate and amend these alternative 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.

[8] 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 4, we think that it is possible for young children to pretend without having *any* beliefs about pretense or other mental states. In those cases, the child might have a conditional belief that guides the pretend behavior, but no beliefs about pretense. Nonetheless, adults and older children clearly do have beliefs about what they are pretending, and they can report on those beliefs. 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. Goldman, 1993; Gopnik, 1993), and in this paper 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, we might exploit the same (or similar) mechanisms to recognize and report on our pretenses (for more details, see Nichols & Stich, in press a,b).

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 4.1 and metarepresentational accounts in Section 4.2.

4.1. 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, this theory maintains, 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. Fig. 4 is a sketch of this process.

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*

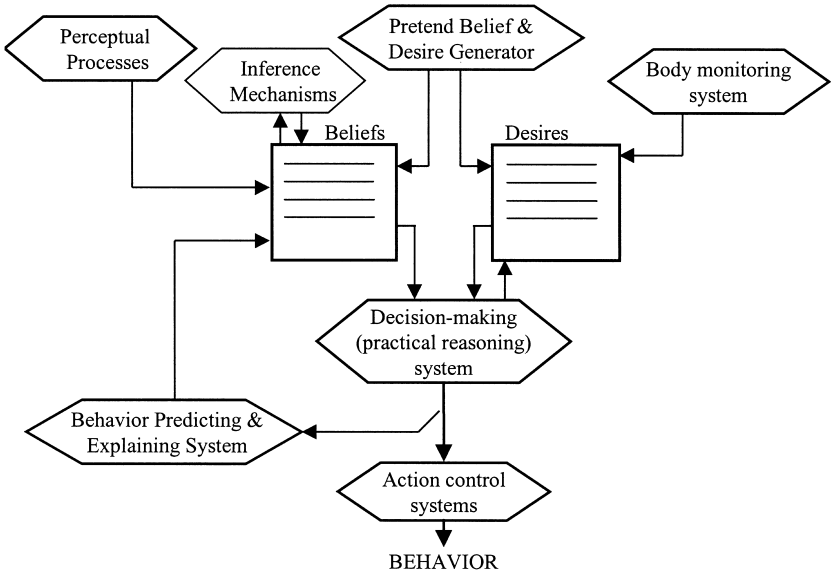

Fig. 4. Off-line simulation account of behavior prediction.

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. Fig. 5 is a rendition of this ''on-line simulation'' account of pretense.

Robert Gordon has been a leader in developing the off-line simulation account of mental state prediction (e.g. Gordon, 1986), 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.[9] ''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

[9] Harris (Harris, 1991, 1995; Harris & Kavanaugh, 1993) and Currie (1995b) present alternative simulation accounts of pretense. Space limitations preclude us from discussing these views here, but see Nichols and Stich (in press b).

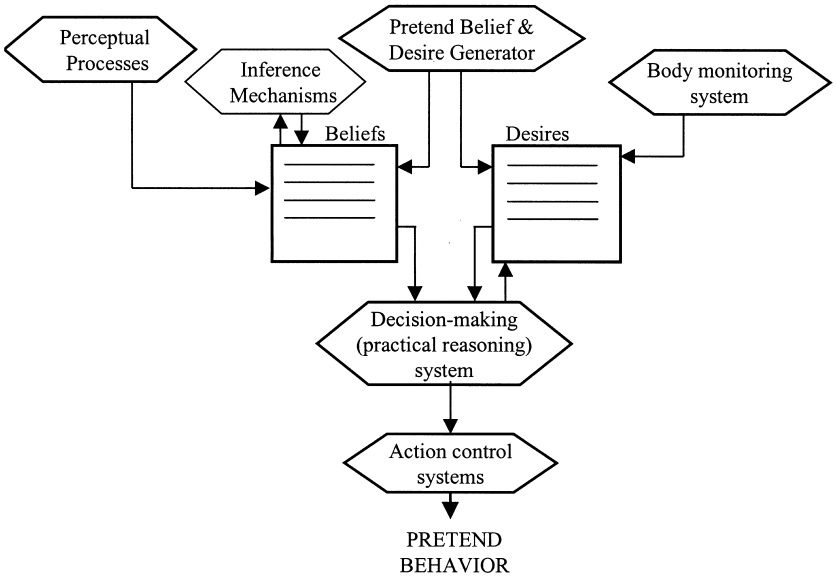

Fig. 5. On-line simulation account of pretense.

> able to obtain answers to questions not addressed in the initial premise (Gordon & Barker, 1994, p. 171).

Two paragraphs later, Gordon and 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 (Gordon & Barker, 1994, p. 172).

Unfortunately, Gordon and Barker do not say much more about the processes subserving pretense, and thus we can't be certain that Fig. 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. In the following section we pose some problems and suggest some revisions to Gordon and Barker's theory. It is our contention that the most promising proposals for patching the Gordon and Barker theory transform their theory into one that is indistinguishable from ours.

4.1.1. Some problems and repairs for the Gordon and Barker theory of pretense

The first problem 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.

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 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.

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.

Since there are independent reasons (set out in Section 3.2) 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. With these extensions to Gordon and Barker's account, the revised theory is growing to look a lot like ours.

A third problem confronting the sort of theory sketched in Fig. 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. But consider 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 is 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. 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.

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, for example, 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. 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) 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 3.3. 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. 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.

As they set it out, Gordon and Barker's account clearly has many shortcomings. They might, of course, accept the various additions and revisions we've proposed. If, as we would hope, they take our suggestions on board as friendly (albeit occasionally quite fundamental) amendments, the resulting account is indistinguishable from the theory that we've proposed.

4.2. 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 & Roth, 1993; Leslie & Thaiss, 1992). 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 3, though the other is not.

A central concern of Leslie's theory is the avoidance of what he calls ''representational abuse'', the cluster of problems that would arise without what we earlier called ''cognitive quarantine''. The infant, Leslie notes, must ''have some way of marking information from pretend contexts to distinguish it from information from serious contexts'' (Leslie, 1987, p. 416). What Leslie proposes 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'' (Leslie, 1987, p. 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). 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 very similar to part of the theory that we have been defending. For, as we noted in Section 4.1.1, to claim that a 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 functional or computational properties, Leslie's hypothesis that representational abuse is avoided because the representations subserving pretense are ''quarantined'' or ''marked off'' (Leslie, 1987, p. 415) is equivalent to claiming, as we do in our theory, that pretense-subserving representations are in a box of their own.[10] 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'' (Leslie, 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, 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'' (Leslie, 1987, p. 424) and thus that ''pretense is an early manifestation of what has been called *theory of mind*'' (Leslie, 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 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'' (Leslie, 1987, p. 416). In developing this idea, Leslie proposes ''a second extension to the primary code'' to work in conjunction with the quotation marks

[10] As we discuss below, Leslie actually disavows the Pretense Box hypothesis. 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, according to which pretense representations have a distinctive content and are stored in the Belief Box. So, on Leslie's account, unlike ours, pretense representations are quarantined from other representations both by their function and by their content.

whose job it is to quarantine the pretense-subserving representations. ''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''**'' (Leslie, 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''** (Leslie, 1987, p. 420). Since Leslie claims that all pretense-representations include the **PRETEND** concept, the pretense box is not needed to distinguish pretense-representations from other mental states. And, indeed, Leslie explicitly rejects the pretense box. Ironically, the clearest place this comes out is in a paper he co-authored with us (Nichols et al., 1996).

> According to Leslie, in terms of boxology, there is no such thing as the 'pretend box', and thus no such thing as simply 'having a pretend'. Instead, pretending is a special case of placing a representation in the 'belief box', where the representation says in effect, 'someone is pretending such and such' (Nichols et al., 1996, p. 56).

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 both 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 3.

Let us start with some facts that are not in dispute. As we noted in Section 3, 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. 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.[11] 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 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, p. 75 for a similar argument.)[12]

Why does Leslie think that pretense is ''a primitive manifestation of the ability to

[11] In a recent paper, Currie (1998) has explored this parallel in some detail, and our development of the point has been significantly influenced by Currie's discussion.

[12] This is not to say that the young child has no understanding of pretense at all. Rather, we think that the young child has what we will call a 'behavioral' understanding of pretense, a notion explained below.

conceptualize mental states'' (Leslie, 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. 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 (pers. commun.; Leslie, 1987, pp. 415–416; Nichols et al., 1996, p. 56). 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 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 a ''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 a ''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 state, viz. pretending that p*. This account is ''mentalistic'' insofar as it invokes the idea that the behavior is produced by underlying mental states of a certain kind (see also Harris, 1994b, p. 251). Now, as Leslie notes, if a child has no understanding at all of pretense, then pretense behavior will often seem utterly bizarre and puzzling (Leslie, 1987, p. 416). (Why on earth would Moma be talking to a banana?!) But by the age of 2 years 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, as several researchers have noted, the child need not have a *mentalistic* understanding of pretense (Harris, 1994b, pp. 250–251; Jarrold, Carruthers, Smith & Boucher, 1994, p. 457; Lillard, 1996, p. 1718). 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 group pretense it does not follow that the child is exhibiting ''a primitive manifestation of the ability to conceptualize mental states''.

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. Although the issue is controversial, 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 (e.g. Jarrold et al., 1994; Lillard, 1993, 1996; Lillard & Flavell, 1992; Rosen, Schwebel & Singer, 1997). 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.'' (Leslie, 1987, p. 416). This, Leslie suggests, is parallel to the fact that a person can pretend that a cup is 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.'' (Leslie, 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 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 of 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, Leslie & Frith, 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, in press b). 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''**.

Where does all of this leave us? As we see it, the arguments for 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''**, are 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. 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 3.

4.2.1. Perner's account

Josef Perner, another leading developmental theorist, has offered an account of pretense that is, in certain respects, rather similar to Leslie's. Perner's account, like Leslie's, is motivated by the fact that in order to avoid confusion, pretend-representations need to be ''quarantined'' from representations of reality. However, in contrast with Leslie, Perner maintains that 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'' (Perner, 1991, p. 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 ''non-real'' 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 (Perner, 1991, pp. 60–61; see also Perner, 1991, pp. 54–56, Perner, 1993, pp. 117–118).

Though Perner's proposal can be interpreted in a variety of ways, we think the best interpretation is that, on Perner's view, pretense representations are distinguished from beliefs in two ways: (i) they have different functional roles, and (ii) they have systematically different meanings or contents. We are encouraged in this reading of Perner by evidence of two quite different sorts. The first is a passage in Perner (1991) in which he maintains that the markers that quarantine pretense from

belief ''modify whether the model represents a *real* or a *hypothetical* situation'' and that they also ''direct 'internal' use and thereby allow differentiation between *real* and *hypothetical*'' (Perner, 1991, p. 35). The second is that Perner himself has maintained that this is the proper interpretation of his view. ''I am…happy to be put in with Alan Leslie as claiming that the MRCs [the ''metarepresentational comments'' 'real' and 'hypothetical'] make both a difference in function and content… My quarrel with [Leslie] concerned only the kind of context marker one uses. I opted for a weaker marker that differentiates only real from hypothetical… Leslie had opted for the stronger marker 'pretend'.'' (pers. commun.)[13]

The problem we have with this view is that it is not at all clear what the difference in content is supposed to be between belief representations and pretense representations. In Perner (1991), it seems that the difference in content is that they ''represent two different situations: the real situation and a hypothetical situation'' (p. 54). However, since it is possible to both pretend and believe that the cup is empty, it is difficult to see how these represent different situations. On the contrary, there seems to be no reason to think that a pretend representation can't have exactly the same content as a belief, in much the same way that a desire can have exactly the same content as a belief. Using a marker to indicate that pretense representations have different functional roles from beliefs is, as we noted earlier, the equivalent of positing a separate box for pretense representations, and that suffices to quarantine pretenses from beliefs. Positing an additional difference at the level of content does no additional work. As a result, pending a further explication of the difference in content between pretense representations and belief representations, we see no reason to adopt the view that there is a systematic difference in the contents of our pretenses and our beliefs.

5. 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 indicate some of the ways in which it differs from other accounts of pretense. 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. The evolutionary function of the PWB, we've suggested, is to enable hypothetical reasoning. Pretense representations on our theory are not distinguished from beliefs in terms of the content of the representations. Here we differ sharply from both Leslie and Perner. In pretense episodes the set of representations being built up in the PWB is inferentially elaborated and updated by the same

[13] We are grateful to Perner for pointing out the passage in Perner (1991) and for allowing us to quote from his written comments on an earlier draft of this paper.

inference and UpDating mechanisms that operate over real beliefs. The importance of the UpDating mechanism in avoiding inferential chaos is another central theme in our theory which sets it apart from other accounts in the literature. In addition to inferential elaboration, pretenders also elaborate the pretense in non-inferential ways, exploiting what we have called the Script Elaborator. One of the virtues of the architectural explicitness of our theory is that it makes clear the need for a Script Elaborator (a point on which other theorists have said relatively little) and it underscores how little we know about how this component of the mind works. All of this cognitive architecture is, we think, essential to both imagination and pretense, but it does not explain why pretenders *do* anything – why they actually enact the pretend scenarios. That is a problem about which a number of leading theorists, including Leslie and Perner have said very little. On our theory, the motivation for pretend play derives from a real desire to act in a way that fits the description being constructed in the PWB. This, we've argued, is a much more satisfactory account than the proposal, hinted at by Gordon and Barker and other simulation theorists, that pretense behavior is motivated by ''pretend desires''. Finally, while our account does not claim that pretense requires mindreading or theory of mind capacities – here we differ sharply with Leslie – the account does leave open the possibility that pretense and mindreading capacities use some of the same mechanisms – the PWB is an obvious candidate here – and thus that breakdowns would be correlated.

While there are obviously many points on which we disagree with other theorists and a number of hitherto neglected issues that our account addresses, it is also the case that our theory is a highly eclectic one which borrows many ideas from other theorists. Our central goal in this paper has been to show how ideas taken from competing and incompatible theories, along with some new ideas of our own, can be woven together into a theory of pretense which is more explicit, more comprehensive and better able to explain the facts than any other available theory. We would, of course, be delighted if other theorists who began from very different starting points were to agree that the eclectic synthesis we have proposed is (near enough) the account toward which they too have been heading.

Acknowledgements

We would like to thank the participants in our study. We would also like to thank Luca Bonatti, Gregory Currie, Alison Gopnik, Paul Harris, Chris Jarrold, Barbara Landau, Alan Leslie, Angeline Lillard, Josef Perner, Richard Samuels, Brian Scholl, Jonathan Weinberg, and an anonymous referee 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., Leslie, A., & Frith, U. (1985). Does the autistic child have a ''theory of mind''? *Cognition*, *21*, 37–46.

Braine, M. (1994). Mental logic and how to discover it. In J. Macnamara, & G. Reyes, *The logical foundations of cognition*, Oxford: Oxford University Press.

Bretherton, I. (1989). Pretense: the form and function of make-believe play. *Developmental Review*, *9*, 383–401.

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

Carruthers, P., & Smith, P. (1996). *Theories of theories of mind*, Cambridge: Cambridge University Press.

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

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

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

Currie, G. (1995c). The moral psychology of fiction. *Australasian Journal of Philosophy*, *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, *Theories of theories of mind*, Cambridge: Cambridge University Press.

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

Currie, G. (1998). Pretence, pretending and metarepresenting. *Mind and Language*, *13*, 35–55.

Davies, M., & Stone, T. (1995a). *Folk psychology: the theory of mind debate*, Oxford: Blackwell.

Davies, M., & Stone, T. (1995b). *Mental simulation: evaluations and applications*, Oxford: Blackwell.

Fein, G. (1981). Pretend play in childhood: an integrative review. *Child Development*, *52*, 1095–1118.

Goldman, A. (1992). Empathy, mind, and morals. *Proceedings and Addresses of the American Philosophical Association*, *66* (3), 17–41.

Goldman, A. (1993). The psychology of folk psychology. *Behavioral and Brain Sciences*, *16*, 15–28.

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, *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, *Natural theories of mind*, Oxford: Blackwell.

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

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

Harris, P. (1994b). Understanding pretense. In C. Lewis, & P. Mitchell, *Children's early understanding of mind: origins and development*, (pp. 235–259). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Harris, P. L., & Kavanaugh, R. D. (1993). Young children's understanding of pretense. Monographs of the Society for Research in Child Development 58(1).

Higginbotham, J. (1995). Tensed thoughts. *Mind and Language*, *10*, 226–249.

Jarrold, C., Carruthers, P., Smith, P., & Boucher, J. (1994). Pretend play: is it metarepresentational? *Mind and 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. (1994). Pretending and believing: issues in the theory of ToMM. *Cognition*, *50*, 211–238.

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

Leslie, A., & Thaiss, L. (1992). Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition*, *43*, 225–251.

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

Lewis, D. (1986). *On the plurality of worlds*, New York: 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 C. Lewis, & P. Mitchell, *Children's early understanding of mind: origins and development*, (pp. 211–234). Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Nichols, S., & Stich, S. (1999a). How to read your own mind: a cognitive theory of self-consciousness. In Q. Smith, & A. Jokic, *Aspects of consciousness*, Oxford: Oxford University Press, in press.

Nichols, S., & Stich, S. (1999b). *Mindreading*, Oxford: Oxford University Press, in press.

Nichols, S., Stich, S., & Leslie, A., et al. (1995). Choice effects and the ineffectiveness of simulation: response to Kuhberger et al. *Mind and Language*, *10* (4), 437–445.

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

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, & D. Olson, *Developing theories of mind*, (pp. 141–172). Cambridge: Cambridge University Press.

Perner, J. (1991). *Understanding the representational mind*, Cambridge, MA: MIT Press.

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

Perner, J., Barker, S., & Hutton, D. (1994). Prelief: the conceptual origins of belief and pretense. In C. Lewis, & P. Mitchell, *Children's early understanding of mind: origins and development*, (pp. 261–286). Hillsdale, NJ: Lawrence Erlbaum Associates.

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*, New York: Norton Translated by C. Gattegno and F. M. Hodgson.

Pylyshyn, Z. (1987). *The robot's dilemma: the frame problem in artificial intelligence*, Norwood, NJ: Ablex.

Rosen, C., Schwebel, D., & Singer, J. (1997). Preschoolers' attributions of mental states in pretense. *Child Development*, *68*, 1133–1142.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding: an inquiry into human knowledge structures*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Sorensen, R. (1998). Self-strengthening empathy. *Philosophy and Phenomenological Research*, *58*, 75–98.

Stich, S., & Nichols, S. (1992). Folk psychology: simulation or tacit theory. *Mind and Language*, *7*, 35–71.

Stich, S., & Nichols, S. (1995). Second thoughts on simulation. In A. Stone, & M. Davies, *Mental simulation: evaluations and applications*, (pp. 87–108). Oxford: Basil Blackwell.

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

Stich, S., & Warfield, T. (1994). *Mental representation*, Oxford: Basil Blackwell.

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*, Cambridge, MA: Harvard University Press.

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