The pretense debate ☆

Stephen Stich *, Joshua Tarzia

Department of Philosophy & Center for Cognitive Science, Rutgers University, 106 Somerset St., 5ᵗʰ Floor, New Brunswick, NJ 098901-4800, USA

ARTICLE INFO

ABSTRACT

In a number of publications, Alan Leslie and colleagues have developed a theory of the psychological mechanisms underlying pretense. This theory maintains that pretense is an early manifestation of "theory of mind" or "mindreading" – the capacity to attribute mental states to oneself and others. Nichols and Stich proposed an alternative theory of pretense on which pretense in young children does not require mindreading. Rather, they argued, young children have a behavioral understanding of pretense. In a lengthy critique, Friedman and Leslie made a persuasive case that the Nichols and Stich theory cannot account for the early emergence of children's capacity to engage in joint pretense and recognize pretense in others. In this paper, we set out a new "pretense game" theory of pretense that avoids the problems raised by Friedman and Leslie, and does not require that children who engage in joint pretense must have a theory of mind. We go on to argue that our pretense game theory can explain many of the facts about pretense that go unexplained in Leslie's theory. The central shortcoming of Leslie's theory is that it attempts to explain the production and recognition of pretense behavior by positing the existence of an innate concept, without explaining *how* this concept enables those who have it to recognize or produce pretense behavior.

1. Introduction

Over the last 25 years, Alan Leslie and his collaborators have developed and defended a sophisticated and influential account of the psychological mechanisms underlying pretense, focusing primarily on pretense in children (German & Leslie, 2001; Leslie, 1987, 1988, 1994, 2002; Leslie & Roth, 1993; Leslie & Thaiss, 1992; Onishi, Baillargeon, & Leslie, 2007). In *Mindreading*, Nichols and Stich (2003) proposed an alternative account that borrowed a number of important ideas from Leslie, while arguing that one central feature of Leslie's account is problematic and should be abandoned. Friedman and Leslie (2007) responded to Nichols and Stich (hereafter N&S), arguing that the N&S account of pretense is fatally flawed because it cannot account for the capacity to recognize pretense – a capacity which emerges quite early in childhood. In this paper, we have a pair of goals, one positive and one negative. The positive goal is to respond to the Friedman and Leslie (hereafter F&L) critique by showing how, with some

elaboration and reconstruction, an account similar to the one proposed by N&S can address the pretense recognition problems posed by F&L. If our revised version of the N&S theory succeeds in meeting F&L's objections, it might be thought that the debate has reached an impasse, at least for the moment, since there are two competing theories that can explain the available facts about pretense. However, our negative goal is to argue that this is not the right conclusion to draw, because the theory of pretense that Leslie and colleagues have developed and defended thus far is not a serious competitor to the theory we propose. More specifically, we will argue that the explanation of pretense that Leslie has proposed is importantly incomplete; it does not provide a satisfying explanation for some of the most obvious and important facts about the production and recognition of pretense.

2. Background: some shared assumptions about cognitive architecture

The N&S account of pretense is set out as a series of additions to a widely shared picture of the basic architecture of the cognitive mind. Though Leslie has explicitly embraced this picture only once, in a paper co-authored with N&S (Nichols, Stich, Leslie, & Klein, 1996), we think it is clear that Leslie, like many other theorists, assumes that this account of cognitive architecture is by and large

correct, though far from complete.[1] So we will begin with a quick overview of the basic cognitive architecture sketched by N&S.[2] The mind, on this account, contains two functionally different kinds of representational states, beliefs and desires. Saying that they are functionally different is a shorthand way of saying that these two kinds of states are caused in different ways and have different patterns of interaction with other components of the mind. Some beliefs are caused by an array of perceptual processes, while others are generated from pre-existing beliefs via a variety of inferential processes. Some desires (like a desire to eat or a desire to be warmer) are caused by systems that monitor states of the body; others, the "instrumental desires," are generated via a process of practical reasoning that takes beliefs and pre-existing desires as input. Still other desires are caused by psychological mechanisms whose nature and function are not well understood. In addition to generating instrumental desires, the practical reasoning system must also determine which desires will be acted on at any given moment. When that decision is made, the information is passed on to an assortment of action-controlling systems which coordinate the behaviors necessary to carry out the decision. Using rectangular "boxes" to represent systems storing functionally similar states, and hexagons to represent psychological processes and mechanisms, N&S offer Fig. 1 as a schematic representation of the basic architecture of the cognitive mind.

Both N&S and Leslie combine this basic architectural picture with a representational account of cognition which maintains that beliefs, desires and other propositional attitudes are representational states. On this account, to have a belief or a desire with the content that p is to have a representation token with that content stored in the Belief Box or the Desire Box. Leslie typically assumes that these representations are linguistic or quasi-linguistic in form, and though N&S take no stand on that issue, for ease of exposition, we will go along with that assumption in what follows.

3. Some points of agreement between Leslie's theory of pretense and the N&S theory

Two examples of pretense recounted by Leslie (1987, 1994) have been widely discussed in the literature. In one of these, a child and her mother pretend that a banana is a telephone. The mother holds the banana up to her face and talks to Daddy. She then hands it to the child, who says hello to Daddy. In the other, a child and the experimenter have a pretend tea party during which imaginary tea is poured into a pair of cups, and then one cup is turned upside down and shaken. The child is asked to point to the full cup and to the empty one. Though in reality both cups have been empty throughout the episode, children as young as two typically have little difficulty identifying the cup that has been turned upside down as the empty one. These examples illustrate a central concern of Leslie's theory of pretense. In both cases, it is plausible to suppose that the child is relying on mental representations that "distort reality" by representing states of affairs that clearly do not obtain. But, Leslie asks, "if a representational system is developing, how can its semantic relations tolerate distortion in these more or less arbitrary ways? Indeed, how is it possible that young children can disregard or distort reality in any way and to any degree at all? Why does pretending not undermine their representational system and bring it crashing down?" (Leslie, 1987, 412) The solution Leslie proposes is that the representations underlying pretense "must somehow be marked off or 'quarantined'" from the

child's "primary representations" – the representations in the child's Belief Box that serve to represent reality as the child actually believes it to be. Though Leslie's primary focus is on pretense in children, it is clear that much the same problem arises for adults engaging in pretense. In the pretend tea party experiment, neither the child nor the experimenter can simply add a representation with the content *the blue cup is full of tea* to the other representations in their Belief Boxes, since they both already have representations with the content *the blue cup is empty*, and the obvious contradiction would presumably lead to untoward consequences for both the pretense and the pretenders.

To deal with this problem, Leslie hypothesizes that the representations that underlie pretense are "marked" in a special way to indicate that their functional role is different from the "primary representations" in the Belief Box. In the terminology Leslie adopts, these marked representations are "decoupled" copies of primary representations which no longer have their "normal input–output relations" or their "normal computational consequences." (1987, 419) The notational device that Leslie uses to mark the decoupled representations underlying pretense is to enclose them in quotation marks. In his early work, Leslie labeled these decoupled representations "metarepresentations," though to avoid confusion and make clear that the label is intended as a technical term in his theory, in his more recent work he often prefers the label "M-representation."

As noted earlier, N&S use "boxes" as a way of representing systems storing functionally similar states. So enclosing pretense representations in quotation marks is simply a notational variant of assigning them to a box of their own. That is the strategy that N&S adopt in their account of pretense. In some of his work focused on the imagination, Nichols (2004) calls this the "pretense box." But since N&S think that representations in this box also play a role in a variety of other cognitive processes, including counterfactual reasoning and third-person mindreading, N&S decided to label this functional component of the mind the "Possible World Box" (or the PWB).[3]

After adding the PWB to their picture of cognitive architecture, N&S offer a detailed account of the role that the PWB plays during an episode of pretense. They explain how many Belief Box representations get added to the PWB, how it manages to avoid contradictions, and how the PWB develops a detailed description of the imaginary world in which the initial premise of the pretense – e.g. *We are having a tea party* – is true (N&S, 2003, §2.4, pp. 28–38). Leslie's theory is silent on all of these issues, and though he does offer a few hints (Friedman and Leslie, 2007, 121; Friedman, Neary, Burnstein, & Leslie, 2010, 318), he does not provide an extended account of how a detailed mental representation of what is going on in an episode of pretense is constructed. This is an important lacunae that will loom large in Section 7, where our goal is to argue that Leslie's theory fails to provide an explanation for some of the most obvious and important facts about pretense. But for the moment, we will put the problem off to the side. The point we want to stress in this section is that both Leslie's theory and N&S's theory recognize the importance of "quarantining" the mental representations describing what is going on in a pretense episode from the mental representations whose job it is to store what the cognitive agent believes about the real world. And, apart from terminological preferences, the two theories are in complete agreement on how to do this.[4]

---

[1] One additional bit of evidence that Leslie accepts something like the account of basic cognitive architecture that N&S set out is that F&L offer no objection to this account, though it is the foundation on which the N&S account of pretense is built.

[2] For more detail, see N&S (2003), Ch. 1.

[3] In their insightful elaboration of the N&S model, Weinberg and Meskin (2006) call it the "imagination box".

[4] As N&S note repeatedly, this is because this part of their theory was heavily influenced by Leslie's pioneering work.