New Ideas in Psychology 40 (2016) 48–64

Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: www.elsevier.com/locate/newideapsych

From two systems to a multi-systems architecture for mindreading

Wayne Christensen [a, *], John Michael [b]

[a] Macquarie University, University Avenue, Macquarie Park, Sydney 2113, Australia
[b] Center for Subjectivity Research, University of Copenhagen, Copenhagen, Denmark

ARTICLE INFO

Article history:
Available online 3 February 2015

Keywords:
Social cognition
Theory of mind
Development
False belief
Representation

ABSTRACT

This paper critically examines Apperly and Butterfill's parallel 'two systems' theory of mindreading and argues instead for a cooperative multi-systems architecture. The minimal mindreading system (system 1) described by Butterfill and Apperly is unable to explain the flexibility of infant belief representation or fast and efficient mindreading in adults, and there are strong reasons for thinking that infant belief representation depends on executive cognition and general semantic memory. We propose that schemas, causal representation and mental models help to explain the representational flexibility of infant mindreading and give an alternative interpretation of evidence that has been taken to show automatic, fast and efficient belief representation in adults.

1. Introduction

In a standard false belief task, a child (or other test subject) observes as an agent places an object (sometimes a doll) in a box at location A and then temporarily departs, whereupon some other agent arrives on the scene and transfers the object to a box at location B. When the first agent returns, the child is asked where the first agent (Sally) is likely to search for the object.[1] The correct answer, of course, is that Sally is likely to search in the box at location A, because that is where she falsely believes the doll to be located (Baron-Cohen, Leslie, & Frith, 1985; Call & Tomasello, 2008; Wimmer & Perner, 1983). This task has been regarded as a litmus test for the capacity to represent the beliefs of other agents because the child can't use her own knowledge of the location of the doll to predict where Sally will search; the child must distinguish Sally's belief about the location of the doll from the actual location. It has been a robust finding that children under the age of four years tend to fail at the standard false belief task, which prompted a widespread view that children younger than about four don't represent beliefs in others. This view has come under pressure, however, from recent evidence that children can succeed on various modified versions of the task that don't require a verbal response by around their first birthdays (e.g. Baillargeon, Scott, & He, 2010; Buttelmann, Carpenter, & Tomasello, 2009; Onishi & Baillargeon, 2005; Song, Onishi, Baillargeon, & Fisher, 2008; Southgate, Senju, & Csibra, 2007; Surian, Caldi, & Sperber, 2007; Träuble, Marinović, & Pauen, 2010), or indeed even by seven months (Kovács, Téglás, & Endress, 2010).

Thus, with evidence of mindreading in early childhood now having been reported by researchers from various labs using diverse methods, the theoretical challenge is to reconcile the large discrepancy in results between verbal false belief tasks and non-verbal versions. As we shall see, this puzzle raises many complex, difficult questions not only about the development of mindreading but also about

* Corresponding author. Tel.: +61 498 109157.
E-mail addresses: wayne.christensen@gmail.com (W. Christensen), johnmichaelaarhus@gmail.com (J. Michael).

[1] The experiments are easier to interpret when described concretely, and in what follows we will refer to the first and second agents as Sally and Anne, and the object as a doll, except where there are variations and the specific details matter.

http://dx.doi.org/10.1016/j.newideapsych.2015.01.003
0732-118X/

the nature of mental representation and cognitive architecture. In this paper we'll examine in some detail one attempt to resolve the puzzle: Apperly and Butterfill's ‘two systems’ account of mindreading (Apperly, 2011; Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). This account addresses many of these issues in insightful ways, providing a characterization of key representational differences between infant and adult mindreading that is linked to an analysis of a tension between efficiency and flexibility, which in turn provides an explanation for why there should be two mindreading systems.

Our discussion identifies a number of points of agreement with Apperly and Butterfill. We think they are right that multiple systems contribute to mindreading, and that some of these systems employ representations that lack the full structure of adult psychological concepts. We also agree that the nature of these systems and their relations is fundamentally shaped by a trade-off between efficiency and flexibility. But we argue for a different architectural solution to these constraints. Apperly and Butterfill believe that the competing demands of efficiency and flexibility give rise to a *parallel two systems architecture* for mindreading. Specifically, they believe that the tension between efficiency and flexibility is such that the only way that both can be achieved in mindreading is by means of distinct cognitive systems. Moreover, because efficiency is attained by means of hard constraints, the efficient system should be capable of only limited information exchange with the flexible system, and the two systems should consequently operate largely in parallel. We'll argue that this is oversimplified, and propose instead a *cooperative multi-system architecture* for mindreading. Efficiency can be compatible with rich information exchange amongst multiple cognitive systems, and a cooperative multi-system organization can yield a better overall balance of efficiency and power than a parallel architecture. Furthermore, recent evidence for infant mindreading suggests that it involves a kind of flexibility that is better explained by a cooperative multi-system architecture.

We also criticize the assumptions that Apperly and Butterfill make concerning the representational characteristics of their two systems. On the one hand, they propose that efficient mindreading is achieved by a simple, inflexible representational scheme, and on the other hand, they propose stringent requirements on the belief representation performed by the flexible mindreading system, which they see as representing beliefs ‘as such’. We argue that the motivations for these proposals are misguided. Specialized cognitive systems can be capable of relatively complex forms of representation, and flexibility can be achieved through cooperative multi-system interactions. We also argue that flexible conceptual belief representation can be simpler and more heterogenous than Apperly and Butterfill suppose. Infants probably do not represent beliefs ‘as such’ in the way that older children and adults do, but it's likely that their belief representations involve generalized semantic memory, and that they are developing forms of conceptual belief representation that serve as a basis for the more sophisticated forms of belief representation that emerge in older children. This learning process probably involves the acquisition of a rich stock of schemas and a constructivist progression in which schemas are refined and more abstract conceptions are developed.

2. Evidence for mindreading in infants

Evidence for the representation of false belief in infants first arose from experiments that used looking behavior instead of verbal response as the measure indicating the presence of the representation. Clements and Perner (1994) found an intriguing inconsistency in young children's responses in a false belief situation: 90% of the children between 35 months and four-and-a-half years old looked first to the empty location (where the Sally falsely believed the object to be located), and yet only 45% of them gave the correct verbal answer when asked where Sally was likely to search. The authors speculated that the children's anticipatory looking might indicate that they implicitly represented Sally's false belief.

This result foreshadowed a new approach to the false belief paradigm focusing on non-verbal measures, especially looking behavior.[2] Onishi and Baillargeon (2005) found that children looked longer when Sally searched in the object's actual location compared with when Sally searched in the original (incorrect) location. According to the authors, the increased looking time indicated a violation of expectation, which revealed that the child had implicitly formed the expectation that Sally would search in the wrong location because she had a false belief. This finding has since been corroborated by numerous groups using similar paradigms: Surian et al. (2007), for example, observed the same pattern in a study involving 13 month-olds; Kovács et al. (2010) found similar evidence in a study with seven-month-olds; and Southgate et al. (2007), measuring children's anticipatory looking as in the Clements and Perner (1994) study, found evidence that 25 month-old children first looked toward the wrong location upon Sally's return (after Sally had briefly looked away and failed to witness that a ball was transferred from one box to another), apparently in anticipation that Sally would search for the ball at the wrong location.

Whether these experiments reveal belief representation has been controversial.[3] But converging results have been reported in recent years using a variety experimental paradigms and measures. Thus, some experiments have probed whether infants are sensitive to variations in the epistemic conditions that produce belief. Senju, Southgate, Snape, Leonard, & Csibra (2011) gave 18-month-olds experience with either an opaque or a transparent “trick” blindfold, then showed them a version of the false-belief scenario in which Sally wore the blindfold that the child had just experienced, and so either could (with the trick blindfold) or could not (with the opaque blindfold) see the

[2] It is common to refer to experimental measures that depend on verbal responses as ‘explicit’ and measures based on behavior as ‘implicit’. We will avoid this, however, because we think that it shouldn't be inferred that cognitive processes are implicit if they manifest in behavioral but not verbal measures. See also Carruthers (2013, p. 145), who makes a similar point.

[3] For discussions see Haith and Benson (1998), Kagan (2008), Müller and Giesbrecht (2008), Heyes (2014), and Ruffman (2014).