

Brief article

Two reasons to abandon the false belief task as a test of theory of mind

Paul Bloom^{a,*}, Tim P. German^b

^a*Department of Psychology, Yale University, New Haven, CT 06520-8205, USA*

^b*University of Essex, Colchester, CO4 3SQ, UK*

Received 5 April 2000; accepted 14 June 2000

Abstract

The false belief task has often been used as a test of theory of mind. We present two reasons to abandon this practice. First, passing the false belief task requires abilities other than theory of mind. Second, theory of mind need not entail the ability to reason about false beliefs. We conclude with an alternative conception of the role of the false belief task. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Cognitive development; False belief task; Theory of mind

1. Introduction

The false belief task is one of the classic methods in the study of development. There have been hundreds of articles and chapters concerning this task, comparing the performance of different populations, exploring how success correlates with other abilities, and modifying and extending the task in various ways. This journal has published 19 such articles, including the one in which the task was first described (Wimmer & Perner, 1983) and the one reporting that autistic individuals show a special impairment on this task, more so than normal and retarded children of the same mental age (Baron-Cohen, Leslie & Frith, 1985). There is little doubt that the false belief task has motivated some of the most exciting research in cognitive development, research that bears in important ways on issues such as nativism and modularity, constructivist theories of conceptual change, and the relationship between self-knowledge and knowledge of others (e.g. see collections of papers in

* Corresponding author. Fax: +1-203-432-7172.

E-mail address: paul.bloom@yale.edu (P. Bloom).

Baron-Cohen, Tager-Flusberg & Cohen, 2000; Carruthers & Smith, 1996; Mitchell & Riggs, 2000).

The ‘standard version’ of the false belief task presents the child with a character, Sally, who leaves a desirable object such as a chocolate in her basket, before departing the scene. In her absence, another character, Anne, removes the object and places it in a box. Children are asked to predict, on Sally’s return to the room, where Sally will look for the object (or, sometimes, where she thinks the object is). Four-year-olds tend to succeed at this task – correctly attributing a false belief to Sally, saying that she will look for the object in the basket – while younger children tend to fail (see Wellman, Cross & Watson, *in press*, for review).¹ This has led many scholars to conclude that children undergo a radical shift in their understanding of the mind. For instance, Gopnik (1993, p. 1) claims that “at about age 4, there is an important developmental shift to a representational model of the mind” and Wimmer and Weichbold (1994, p. 45) state that “...not until the age of about 4 years do children become able to attribute belief states to themselves and other people” (see also Flavell, 1988, p. 247).

Under this view, failure at false belief task reflects some serious deficit in children’s understanding of the mental lives of themselves and others – a deficit in ‘theory of mind’.² Our goal here is to review the evidence to the contrary; we argue that there are two independent reasons to abandon the false belief task as a test of theory of mind.

2. Reason 1: There is more to passing the false belief task than theory of mind

The logic behind the false belief task was outlined by several commentators on the paper by Premack and Woodruff (1978) entitled ‘Does the chimpanzee have a theory of mind?’ Suppose you want to know whether a chimpanzee can reason about the mental states of others. As Dennett (1978) and others pointed out, it is not enough to demonstrate that individual A can predict the actions of individual B. In many cases, A can do so without an understanding of B’s mental states, but by simply observing the actual state of the world. (Suppose A knows the chocolate is in the basket and observes B searching for food. A might expect B to look in the basket, not because A

¹ There are other tasks that assess an individual’s ability to attribute a false belief to him or herself. These typically take the form of showing children a familiar container (e.g. a ‘smarties’ tube), soliciting their beliefs about its contents, and then revealing that the container has a different content (e.g. a pencil). They are then asked to state what they thought was in the container (e.g. Perner, Leekham & Wimmer, 1987, among many others), and younger children typically fail to acknowledge their own false belief. In this paper our arguments focus primarily on the Sally–Anne task, but these apply with equal force to this somewhat different task.

² Note that there is a more subtle proposal articulated by Perner (1991, 1995), which states that 2- and 3-year-olds “understand mental states as relations of semantically evaluable objects (propositions)” but that they fail to appreciate “that these intentional objects (propositions) are semantically evaluable” (Perner, 1995, p. 243). The idea here is that children need to learn that beliefs are mental representations that (critically) can *misrepresent* how the world is; this understanding is presumed to develop as part of children’s acquisition of the more general concept of representation. Perner’s theory differs in important regards from the more standard interpretation of the false belief findings; it is considered at greater length elsewhere (German & Leslie, 2000; Leslie, *in press*), and will not be discussed here.

is attributing a belief to B, but because the chocolate actually is in the basket.) A more robust test involves predicting the behavior of another animal based on an inferred mental state that differs from reality – a false belief. This would show that the individual understands that it is the mental state, rather than the state of the world, that causes the action.

Based on this reasoning, Wimmer and Perner (1983) developed a false belief task for use with preschool and school-age children. This task was modified by Baron-Cohen et al. (1985), in order to be simple enough to be administered to children with autism. The Baron-Cohen et al. (1985) version has become the ‘standard’ version of the false belief task, and this is the Sally–Anne task discussed above.

As many have pointed out, even if children understand that beliefs can be false, this is still a difficult task. To solve it, the child has to follow the actions of two characters in a narrative, has to appreciate that Sally could not have observed the switching of the chocolate, has to remember both where the chocolate used to be and where it is at the time of the test, and has to appreciate the precise meaning of the question (for instance, that it means where will Sally look, not where she *should* look). Though control questions ensure that 3-year-old children can cope with some of these basic task demands (e.g. see Perner, Leekham & Wimmer, 1987), the task is too hard for 1- and 2-year-olds, as they lack sufficient attentional and linguistic resources to cope.

Why do 3-year-olds fail? Several investigators have modified the false belief task so as to make it simpler – for instance, by making the questions simpler, more specific, and more pragmatically natural (Freeman, Lewis & Doherty, 1991; Lewis & Osborne, 1990; Moses, 1993; Siegal & Beattie, 1991; Surian & Leslie, 1999), making the change of location less salient (Carlson, Moses & Hix, 1998; Wellman & Bartsch, 1988; Zaitchik, 1991), giving the children a memory aid for false belief content (Freeman & Lacohee, 1995; German & Leslie, 2000; Mitchell & Lacohee, 1991) and so on. Such modified false belief tasks are often passed by 3-year-olds, a finding that has been used to support the argument that younger children have sophisticated conceptual competence when it comes to understanding that beliefs can be false, but that this competence is filtered through inefficient processing capacities (German & Leslie, 2000).

Advocates of the developmental change view are not convinced. Gopnik (1993, p. 13), for instance, sees such findings as showing that “At best... there may be evidence of some fragile and fragmentary false-belief understanding in some 3-year-olds under some conditions.” And, as noted above, such modifications might lower the age of children who pass the task by about 6 months to a year, but not much lower. Because of this, the results above do not in themselves challenge the substance of the developmental change view supported by Gopnik and others, because an option for developmental change theorists is to argue that they merely show that developmental change occurs somewhat earlier than expected.

The more serious problem is that false belief tasks are inherently difficult. This is because any false belief task requires, at minimum, that the child reasons about a belief that is false. As Leslie, among others, has pointed out, beliefs are *supposed to be true*. This is *what they are for* (Leslie, 1994). Hence, even for a child who clearly understands that beliefs can be false, getting the right answer places non-trivial processing demands.

To put it another way, to succeed at the false belief task, the child has to override useful and simple heuristics, such as ‘people will act in accord with their desires’ (Fodor, 1992; Gergely, Nádasdy, Csibra & Biró, 1995; Roth & Leslie, 1998; Russell, Mauthner, Sharpe & Tidswell, 1991; Saltmarsh, Mitchell & Robinson, 1995; Zaitchik, 1990; but see Wimmer & Weichbold, 1994 for an opposing view).

Furthermore, there is evidence that standard false belief tasks are difficult for children even independent of the requirement to reason about false belief. In the ‘false photograph’ task, a picture is taken of a scene, the scene is changed, and the child is asked what the picture depicts. Here there are multiple representations and there is the realist pull of the way the world really is – but there are no false beliefs. If children pass the false photograph task and fail the false belief task, it would be reasonable to infer that their problem with the false belief task really does have to do with beliefs. But in fact, 3-year-olds fail both the false photograph task and the false belief task (Zaitchik, 1990; see also Leslie, 2000; Slaughter, 1998).

It gets worse, as there is evidence that standard false belief tasks are difficult for children, independent of the requirement to reason about *any* sort of representation. For instance, Riggs, Peterson, Robinson and Mitchell (1998) constructed a task in which children are required to set aside an actual state of affairs (the fact that an object has been placed in one location) and generate an answer based on an alternative counterfactual state of affairs (where it would have been placed had some other event not happened). Younger children are poor at this task, even though there is no representational content. In a similar vein, Roth and Leslie (1998) developed the ‘screens task’. Children are presented with a box and a basket, and a marble is placed in the box. These objects are then concealed by a screen. Then a replica box and basket are placed in front of the screen. A marble is placed in the replica box, and then moved to the replica basket. Children are simply asked to report where the marble is behind the screen. This task requires children to set aside a currently available, salient situation (in which the marble is in the basket) and generate an answer based on a similar, less salient situation (in which the marble is in the box). Once again, 3-year-olds fail this task.

It is revealing that when older autistic children are tested on these photograph and screens tasks, they pass – even though they fail on the standard false belief task (Charman & Baron-Cohen, 1995; Leslie & Thaiss, 1992; Roth & Leslie, 1998). This suggests that autistic children really do have specific problems with false beliefs. But given their failure at all of these tasks, this inference cannot be made about normal 3-year-olds.

3. Reason 2: There is more to theory of mind than passing the false belief task

If younger children do not show signs that they have an appreciation of false belief, how do you know that they are reasoning about mental states at all? Fortunately, there are other methods that can be used to determine whether or not children are attributing mental states. For instance, in an elegant study by O’Neill (1996), 2-year-olds observed as an attractive toy was put on a high shelf. As this happened, the child’s parent was either present or absent. When later asking for help in retrieving

the toy, the children were more likely to name the toy and gesture to the location when their parent had not been present to witness the placement of the toy than if their parent had been present. This suggests that they modify their behavior according to the knowledge states of other people (i.e. whether or not their parent possesses a given belief), and that they have a tacit appreciation of the circumstances under which beliefs are formed.

Furthermore, there are plenty of signs that even before their second birthday, children have some appreciation of the workings of other minds. They are capable of initiating pretend play and of understanding the pretence of others (Leslie, 1994). They can attribute goals to other agents (Csibra, Gergely, Biró, Koós & Brockbank, 1999; Gergely et al., 1995; Woodward, 1998). They can imitate the intended, as well as completed actions of other agents (Carpenter, Akhtar & Tomasello, 1998; Meltzoff, 1995). They can use eye gaze as a cue to what someone is attending to when they use a new word (Baldwin, 1991) and orient to follow the gaze of an inanimate object if it displays evidence of being an intentional agent (Johnson, Slaughter & Carey, 1998). In fact, young children's learning of the meanings of words can be seen in a large part as the direct consequence of their ability to infer the referential intentions of other people (Bloom, 2000). It might be that developmental psychologists are so obsessed with the false beliefs task just because it is the one measure of theory of mind that children are not very good at.

Consider finally the fact that older autistic children fail the false belief task. Wimmer and Weichbold (1994) suggest that this "speak[s] for the validity of the false belief task, since social and communicative impairments are among the defining characteristics of autism". But it also illustrates a weakness in the task. Normal 3-year-olds and older children with autism both fail the false belief task, but, in all interesting regards, *normal 3-year-olds are nothing like older children with autism* (e.g. see Happé, 1996). Normal 3-year-olds are far superior with regard to communicative and linguistic skills, the ability to pretend and understand the pretence of others, and the ability to engage in, understand and manipulate the actions of others. This is a severe problem for any theory that lumps the two groups together as individuals who lack theory of mind.

A more promising analysis is that some individuals with autism fail the false belief task because they lack the capacity to acquire a theory of mind. In contrast, 3-year-olds might fail the false belief task because of general task demands, because they don't have a grasp of false belief, or both. But they surely have a 'theory of mind', in the general sense of having a sophisticated ability to reason about the mental states; this is precisely why they differ from autistic individuals in the social, communicative, and imaginative domains.

4. The proper role of the false belief task

The false belief task should not be abolished. For one thing, while failure in the false belief task isn't necessarily informative about a child or animal's conceptual abilities, success is. If a 2-year-old or a chimpanzee could be shown to succeed at a

false belief task, it would be big news. Furthermore, the false belief task can be used to explore the relative difficulty of reasoning about different representations, including beliefs, photographs, and drawings. It can be used to diagnose and study older children and adults with cognitive and linguistic impairments. And studies of the factors that improve or diminish performance on the task (such as manipulations of salience, or use of memory aids) can be interesting in their own right, as they bear on theories of how children cope with multiple representations, how they understand questions, and so on. The point of this paper is that the false belief task should be considered in its proper context. It is an ingenious, but very difficult task that taps one aspect of people's understanding of the minds of others. Nothing more, nothing less.

References

- Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, 62, 875–890.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition*, 21, 37–46.
- Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. J. (2000). *Understanding other minds: perspectives from developmental cognitive neuroscience* (2nd ed.). Oxford: Oxford University Press.
- Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
- Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development*, 69, 672–691.
- Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behaviour and Development*, 21, 315–330.
- Carruthers, P., & Smith, P. (1996). *Theories of theories of mind*. Cambridge: Cambridge University Press.
- Charman, T., & Baron-Cohen, S. (1995). Understanding photos, models, and beliefs: a test of the modularity thesis of theory of mind. *Cognitive Development*, 10, 287–298.
- Csibra, G., Gergely, G., Biró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, 72, 237–267.
- Dennett, D. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 1, 568–570.
- Flavell, J. H. (1988). The development of children's knowledge about the mind: from cognitive connections to mental representations. In J. Astington, P. Harris, & D. R. Olson (Eds.), *Developing theories of mind*. Cambridge: Cambridge University Press.
- Fodor, J. (1992). A theory of the child's theory of mind. *Cognition*, 44, 283–296.
- Freeman, N., & Lacoche, H. (1995). Making explicit 3-year-old's implicit competence with their own false beliefs. *Cognition*, 56, 31–60.
- Freeman, N. H., Lewis, C., & Doherty, M. J. (1991). Preschoolers' grasp of a desire for knowledge in false-belief prediction: practical intelligence and verbal report. *British Journal of Developmental Psychology*, 9, 7–31.
- Gergely, G., Nádasdy, Z., Csibra, G., & Biró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56, 165–193.
- German, T. P., & Leslie, A. M. (2000). Attending to and learning about mental states. In P. Mitchell, & K. Riggs (Eds.), *Children's reasoning and the mind*. Hove: Psychology Press.
- Gopnik, A. (1993). How we know our own minds: the illusion of first person knowledge of intentionality. *Behavioral and Brain Sciences*, 16, 1–14.
- Happé, F. (1996). *Autism: an introduction to psychological theory*. Cambridge, MA: Harvard University Press.
- Johnson, S., Slaughter, V., & Carey, S. (1998). Whose gaze will infants follow? The elicitation of gaze following in 12-month-olds. *Developmental Science*, 1, 233–238.

- Leslie, A. M. (1994). Pretending and believing: issues in the theory of ToMM. *Cognition*, 50, 193–200.
- Leslie, A. M. (2000). How to acquire a 'representational theory of mind'. In D. Sperber, & S. Davies (Eds.), *Metarepresentation*. Oxford: Oxford University Press (in press).
- Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition*, 43, 225–251.
- Lewis, C., & Osborne, A. (1990). Three-year-olds' problems with false belief: conceptual deficit or linguistic artifact?. *Child Development*, 61, 1514–1519.
- Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31, 838–850.
- Mitchell, P., & Lacohée, H. (1991). Children's early understanding of false belief. *Cognition*, 39, 107–127.
- Mitchell, P., & Riggs, K. (2000). *Children's reasoning and the mind*. Hove: Psychology Press.
- Moses, L. J. (1993). Young children's understanding of belief constraints on intention. *Cognitive Development*, 8, 1–25.
- O'Neill, D. K. (1996). Two-year-old children's sensitivity to parent's knowledge state when making requests. *Child Development*, 67, 659–677.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Perner, J. (1995). The many faces of belief: reflections on Fodor's and the child's theory of mind. *Cognition*, 57, 241–269.
- Perner, J., Leekham, S., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: the case for conceptual deficit. *British Journal of Developmental Psychology*, 5, 125–137.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 512–526.
- Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, 13, 73–91.
- Roth, D., & Leslie, A. M. (1998). Solving belief problems: towards a task analysis. *Cognition*, 66, 1–31.
- Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology*, 9, 331–349.
- Saltmarsh, R., Mitchell, P., & Robinson, E. (1995). Realism and children's early grasp of mental representation: belief-based judgements in the state change task. *Cognition*, 57, 297–325.
- Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition*, 38, 1–12.
- Slaughter, V. (1998). Children's understanding of pictorial and mental representations. *Child Development*, 69, 321–332.
- Surian, L., & Leslie, A. M. (1999). Competence and performance in false belief understanding: a comparison of autistic and three-year-old children. *British Journal of Developmental Psychology*, 17, 141–155.
- Wellman, H., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, 30, 239–277.
- Wellman, H. M., Cross, D., & Watson, J. (in press). Meta-analysis of theory of mind development: the truth about false-belief. *Child Development*.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and the containing function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.
- Wimmer, H., & Weichbold, V. (1994). Children's theory of mind: Fodor's heuristics examined. *Cognition*, 53, 45–57.
- Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69, 1–34.
- Zaitchik, D. (1990). When representations conflict with reality: the preschooler's problem with false beliefs and "false" photographs. *Cognition*, 35, 41–69.
- Zaitchik, D. (1991). Is only seeing really believing? Sources of true belief in the false belief task. *Cognitive Development*, 6, 91–103.