Mental files and belief: A cognitive theory of how children represent belief and its intensionality

CrossMark

Josef Perner [a],[*], Michael Huemer [a], Brian Leahy [b]

[a] *Center for Cognitive Neuroscience, University of Salzburg, 5020 Salzburg, Austria*
[b] *Department of Linguistics, University of Konstanz, 78457 Konstanz, Germany*

ABSTRACT

We provide a cognitive analysis of how children represent belief using mental files. We explain why children who pass the false belief test are not aware of the intensionality of belief. Fifty-one 3½- to 7-year old children were familiarized with a dual object, e.g., a ball that rattles and is described as a rattle. They observed how a puppet agent witnessed the ball being put into box 1. In the agent's absence the *ball* was taken from box 1, the child was reminded of it being a *rattle*, and emphasising its being a *rattle* it was put back into box 1. Then the agent returned, the object was hidden in the experimenter's hands and removed from box 1, described as a "rattle," and transferred to box 2. Children who passed false belief had no problem saying where the puppet would look for the *ball*. However, in a different condition in which the agent was also shown that the ball was a rattle they erroneously said that the agent would look for the ball in box 1, ignoring the agent's knowledge of the identity of *rattle* and *ball*. Their problems cease with their mastery of second-order beliefs (she thinks she knows). Problems also vanish when the ball is described not as a *rattle* but as a *thing that rattles*. We describe how our theory can account for these data as well as all other relevant data in the literature.

0. Introduction

We present a theory of how children represent belief based on mental files. Mental files play an important role in philosophy, addressing longstanding issues about Russell's (1910) problem of acquaintance (Recanati, 2012) and Frege's (1892) foundational problems of logics about identity and the sense-reference distinction (Perry, 2002). As discourse referents they play a role in linguistics (Heim, 1982; Kamp & Reyle, 1993; Karttunen, 1976). In psychology they have only been used in isolated places for object files (Pylyshyn, 2007; Treisman & Gelade, 1980). They have not played any significant role in the popular area of "theory of mind" research, apart from some attempts to apply discourse referents to developmental phenomena hidden in the context of philosophy journals (Perner & Brandl, 2005; Perner, Rendl, & Garnham, 2007). This is somewhat surprising since mental files theory in philosophy has been used extensively to deal with the pernicious logical problems created by statements about beliefs and other mental terms (Recanati, 2012).

We intend to change this picture. We provide a coherent theory of belief representation and test this theory with data from children's understanding of belief and its intensionality. In Section 1 we introduce the empirical problems to which we apply our theory. In Section 2 we use this concrete material to introduce our theory. In Section 3 we describe unique predictions of our theory and present the results of testing those predictions.

1. Children understanding intensionality

There is a curious window in child development, which opens when a child first passes verbal false belief tasks[1] around 4 years and closes 2 years later when she passes second order belief tasks. During this period children appreciate that others can have beliefs that differ from their own, e.g.: Mistaken Max did not witness the transfer of his chocolate to a new location and thinks it is still in its original place (Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983). Yet they seem to misunderstand the intensionality

---

* Corresponding author at: Center for Cognitive Neuroscience, University of Salzburg, Hellbrunnerstrasse 34, 5020 Salzburg, Austria.

E-mail addresses: josef.perner@sbg.ac.at (J. Perner), michael.huemer2@stud.sbg.ac.at (M. Huemer), brian.leahy@uni-konstanz.de (B. Leahy).

[1] A false belief task is *verbal* just in case the subject is asked an explicit verbal question about the protagonist's belief or belief-based actions, which requires children to make a judgement about what is the case or what will soon happen. Responses may be made by simple pointing; the subject need not verbalise anything on a verbal false belief task.

of belief, i.e., that belief about an object depends on the label under which the object is known to the believer. Understanding of intensionality appears with passing second-order belief tasks. Our theory can explain these existing data but makes a bizarre prediction of a sharp decline in performance on a novel intensionality task during this window, and subsequent sharp improvement when the window closes. These predictions were tested and confirmed.

This developmental window was discovered by Apperly and Robinson (1998, 2001, 2003). Inspired by Russell (1987), they familiarized 4- to 6-year-old subjects with two objects: a standard eraser and an eraser that was also a die. Then a puppet, Heinz, appeared and saw the two objects but was not informed of the dual nature of the die-eraser. Then the *knowledge question* was posed: "Does Heinz know that the die is an eraser?". Subjects who passed the false belief task had no problem answering "No". But when the *where-look question* was posed, "Where will Heinz look for an eraser?", the same children chose at random between the location of the standard eraser and the location of the die-eraser. Sprung, Perner, and Mitchell (2007) showed that children who passed second-order belief tasks only indicated the location of the standard eraser.

We refer to children who pass both first and second order verbal false belief tasks as (++); children who pass first-order but fail second-order verbal false belief tasks (+−); and children who fail both as (−−). The (+−) children are those in Apperly and Robinson's window. Even though (+−) children deny that Heinz knows that the die is an eraser, they treat this fact as relevant to his behaviour. So their grasp of how mental states determine behaviour seems incoherent. They do not fully understand the intensionality of belief. From "Heinz knows that there is a die in location l" and "The die is an eraser" they infer that Heinz will behave as though there is an eraser in location l. At the same time, they deny that Heinz knows that the die is an eraser. This incoherence is in dire need of explanation. No existing explanation of this incoherence covers all existing data.

Correct answers to the knowledge question might somewhat antedate passing verbal false belief tasks because past evidence has shown that children answer knowledge questions before they understand false belief (Hogrefe, Wimmer, & Perner, 1986; Sodian, Thoermer, & Dietrich, 2006). They should nevertheless be correlated because both tasks require that a subject understands how the information available from an agent's perspective differs from the information available from the child's own perspective. Children must appreciate that from Heinz' perspective, the information that the die is an eraser is unavailable. To predict when children cease to assume that Heinz will go for an eraser to where he knows the die is, Sprung et al. (2007) drew on claims from Clark (1997) and from Tomasello (1999, 2014) that describing an object using different labels provides different *perspectives* on that object. So successfully answering the question about where Heinz will look for an eraser (the where-look question) requires understanding embedded perspectives. Subjects must appreciate that from Heinz' perspective, the eraser-perspective on the die is not available. Thus correctly answering the where-look question is predicted to co-occur with passing tests of second order mental states, where a subject must evaluate whether an agent *thinks* he *knows* something (Perner & Howes, 1992).

Sprung et al. (2007) therefore predicted that 4- to 6-year old children would have no problems with the where-look question if information about an object was provided in a *predicative* manner rather than in an *individuating* manner. To say that an object is a stick is to individuate it in a certain way, to use a label that provides a perspective. To say that a stick is long is to predicate something of it, which does not provide a perspective on the object but rather provides information about it relative to the established perspective of being a stick. In Apperly and Robinson's task, the subject and the agent (Heinz) could both see an eraser and a die

but only the subject knew that the die was also an eraser. In a parallel task in Sprung et al. (2007), the child and the agent could see both a long stick and a second, partially occluded stick, but only the child knew that the occluded stick was also long. In this new task, 4- to 6-year-olds had no problem with either the knowledge question ("Does Heinz know that this stick is long?") or the where-look question ("Where will Heinz go to get a long [stick]?").

Rakoczy, Bergfeld, Schwarz, and Fizke (2015) demonstrate that the curious problem with the where-look question about a die/eraser also disappears when the original task is simplified in that only a single object (the die/eraser) is used without the plain eraser. They conclude that (+−) children understand intensionality; Apperly and Robinson's observations are to be explained by excessive demands on memory load and ambiguity resolution that children cannot meet before they are around 6 years old. The crucial task (their experiment 3) involved a ball[2]/rattle instead of the die/eraser and went as follows. Children were shown that the ball was also a rattle. Then puppet Susi appeared. Child and Susi observed the ball being placed in box 1. Susi left and the ball was removed from box 1. Children were reminded that the ball was also a rattle. The object was returned to box 1. Susi returned. The experimenter reached into the box and made a *hidden transfer*: she hid the object in her hands, removed it, called it a *rattle*, shook it to make it rattle, and moved the rattle to box 2. Subjects were then asked where Susi would look for the *ball*. The correct answer is box 1, since Susi does not know that the rattle is the ball. To pass this test, subjects must suppress their knowledge that the rattle is the ball. Consequently, Rakoczy et al. (2015) maintain that this task tests the same abilities as Apperly and Robinson's first task, where children must suppress their knowledge that the die is an eraser. Rakoczy et al. found that (+−) children had no problems with this task; there was no evidence of Apperly and Robinson's curious window.

We will shortly criticize Rakoczy et al.'s (2015) conclusions. But whether or not the conclusions are right, the data pose a problem for Apperly and Robinson's theory. For in order to pass this test, subjects must be aware that from the puppet's perspective, the ball-perspective on the rattle is not available. Thus Apperly and Robinson would incorrectly predict that success on this task occurs after success on first-order verbal false belief tasks, counter to the findings in Rakoczy et al. (2015). For the same reasons these data also pose a problem for the appeal to embedded perspectives in Sprung et al. (2007).

However, Rakoczy et al. (2015) still have no explanation for why the curious window appears when information about an object is provided in an individuating manner, but disappears when information is provided in a predicative manner. The predicative cases seem to pose the same demands on working memory and ambiguity resolution, yet the effect is not observed. Moreover, Apperly and Robinson (2003) contrasted a false belief condition with a dual identity condition, both of which had only one critical object and identical test questions. They still found a difference in difficulty, so the number of critical objects does not seem to be the source of the difficulty.

In order to account for this inexplicable set of data we now present a mental file theory of belief representation and apply it to this set of data. In addition, we draw new predictions from it and then test these predictions on 3½- to 7-year old children.

2. Mental files

A mental file is a tool for managing information about an object in the world (say, the file's *referent* or *external referent*). Files cap-

---

[2] Rakoczy et al. used a pen/rattle. In our studies we used a ball/rattle, which we use here as our paradigmatic example to make it consistent with later descriptions of the same experimental conditions.