Cognition 173 (2018) 60–74

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Children exhibit different performance patterns in explicit and implicit theory of mind tasks

Nese Oktay-Gür, Alexandra Schulz, Hannes Rakoczy*

Institute of Psychology, University of Göttingen, Germany

ARTICLE INFO

Keywords:
Theory of mind
Two-systems theory
Signature limits
Aspectuality

ABSTRACT

Three studies tested scope and limits of children's implicit and explicit theory of mind. In Studies 1 and 2, three- to six-year-olds (N = 84) were presented with closely matched explicit false belief tasks that differed in whether or not they required an understanding of aspectuality. Results revealed that children performed equally well in the different tasks, and performance was strongly correlated. Study 3 tested two-year-olds (N = 81) in implicit interactive versions of these tasks and found evidence for dis-unity: children performed competently only in those tasks that did not require an understanding of aspectuality. Taken together, the present findings suggest that early implicit and later explicit theory of mind tasks may tap different forms of cognitive capacities.

1. Introduction

One of the biggest puzzles in recent theory of mind (ToM) research is this: how can we reconcile decades of findings that children fail explicit verbal false belief (FB) and related ToM tasks before age 4 with a growing body of evidence that even infants can perform successfully in implicit versions of such tasks?

1.1. The puzzle

In standard verbal FB tasks children are required to make explicit predictions of a protagonist's action on the basis of her mistaken belief. In change-of-location scenarios, for example, the child witnesses a protagonist put an object into box 1. In the protagonist's absence, the object is then transferred to box 2, and the child has to predict where the protagonist will search for it. Children younger than 4 years of age tend to fail in this task by claiming that the protagonist will look in box 2 while older children pass by predicting that she will mistakenly search in box 1 (Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983).

Less explicit versions of such tasks, using looking and non-verbal interactive behavior as dependent measures have revealed competence much earlier than age 4. In violation-of-expectation looking time tasks infants look longer at an event if a protagonist performs an action which does not fit with her (false) belief (e.g., Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007; see Baillargeon, Scott, & Bian, 2016; Baillargeon, Scott, & He, 2010 for review). Furthermore anticipatory looking studies show that two-year-olds form an expectation about the behavior of an agent based on her (false) belief (Clements & Perner, 1994; Southgate, Senju, & Csibra, 2007). Studies using helping behavior have revealed that infants and toddlers spontaneously help their interaction partners in ways that suggest that they are sensitive to the partners' beliefs (Buttelmann, Carpenter, & Tomasello, 2009; Buttelmann, Over, Carpenter, & Tomasello, 2014; Knudsen & Liszkowski, 2012a, 2012b; Southgate, Chevallier, & Csibra, 2010).

1.2. Theoretical responses to the puzzle: early competence versus conceptual change

How, then, might these two sets of findings be theoretically reconciled? From the point of view of early competence accounts (often nativist in spirit), the new findings with implicit measures suggest that the core competence for belief ascription operates from early on, is perhaps even innate, and does not itself undergo fundamental qualitative changes (Baillargeon et al., 2010; Carruthers, 2013; Leslie, 2005). The fact that children fail standard verbal FB tasks until several years later does not reflect, according to such views, any lack of conceptual competence or any significant conceptual development (the conceptual apparatus for belief ascription is present early and thus does not need to undergo substantial development). Rather, standard verbal tasks pose a number of extraneous task demands (in terms of inhibition, linguistic proficiency etc.) and thus mask children's early competence. Such tasks are then only mastered once children have acquired the requisite yet extraneous capacities (executive function etc.) required to meet these task demands (Carruthers, 2013; Leslie, 2005).

Conceptual change accounts, in contrast, assume that there may be

* Corresponding author at: University of Göttingen, Institute of Psychology, Waldweg 26, 37073 Göttingen, Germany.
E-mail address: hannes.rakoczy@psych.uni-goettingen.de (H. Rakoczy).

https://doi.org/10.1016/j.cognition.2018.01.001
Received 16 January 2017; Received in revised form 19 December 2017; Accepted 3 January 2018

different forms and levels of conceptual capacities, with some crucial qualitative conceptual change from infancy onwards. The two kinds of tasks (early implicit and later explicit) may actually not tap the very same kinds of conceptual abilities. Rather, the implicit tasks may tap a more basic form of ToM that develops earlier and may constitute a foundation for the fully-fledged ToM capacities developing subsequently (e.g., Gergely & Csibra, 2003; Perner, 1991; Perner & Roessler, 2012; Wellman, 2011). If such a general picture was accurate, there should be clear differences between the scope and limits of the earlier and more basic, compared to the later-developing and more complex capacities. In particular, there should be signature limits of the early-developing capacities: Agents operating only on the basis of these more basic capacities should be able to master only a sub-set of simpler ToM tasks while failing more complex ones. A recent two-systems-account, in particular, makes clear, theoretically motivated and testable predictions of specific signature limits (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). According to this account, there are at least two systems for tracking mental states – in analogy, for example, to the widely shared assumption that in the domain of numerical cognition there are at least two systems for tracking numbers (e.g. Carey, 2009; Feigenson, Dehaene, & Spelke, 2004). The capacities tapped in implicit tasks reflect the workings of a simpler, evolutionarily and developmentally more ancient system (System 1, S1) that operates fast and independently of central cognitive resources (such as language or executive function). In contrast, the capacities tapped in explicit tasks reflect the workings of System 2 (S2) that develops later, is dependent on language and executive function, and operates with the fully-fledged conceptual grasp of subjective mental (mis-)representation. More specifically, the two systems differ crucially in their representational capacities in the following ways: S1 enables a subject to track so-called *relational attitudes*. These are relations that agents hold to situations like registering (in the sense of: being in perceptual contact with) an event. S1 allows an agent to represent, for example, that from his vantage point Peter can see (is in perceptual contact with) the cake on the table whereas Paul and Mary, from their perspective, cannot. On the basis of S1, an agent can thus engage in level-I perspective-taking. Importantly, though, keeping tracking of what another agent has or has not registered does not yet involve understanding a crucial form of the subjectivity of mental representation, namely their so-called *aspectuality*. Mental (and linguistic) representations are aspectual in the sense that agents represent objects (e.g., Clark Kent, who in fact is Superman) and situations always only under some aspects (e.g., "Clark Kent") and not under others (e.g., "Superman"). Imagine, for illustration, that Peter (ignorant of the Clark Kent = Superman identity) witnesses the following sequence of events: First, he sees Clark Kent enter the house; then he sees Superman exit the house and fly to the beach. In order to understand what Peter believes about Clark Kent's whereabouts, we need to take into account how he has represented the events in this sequence. De facto, he has seen Clark Kent first enter and then leave the house. But crucially, he only saw the person entering the house *as* Clark Kent. The person leaving was not represented as Clark Kent, but only as Superman. In consequence, Peter believes that Clark Kent must still be in the house.

S1, with its restriction to the representation of relational attitudes such as registering an event, does not enable agents to make such fine-grained distinctions regarding the question under which aspects an agent has encountered an object. If an agent has registered Clark Kent leaving the house, she has ipso fact registered Superman leaving the house. Registration is not aspectual. S2, in contrast, recruits fully-fledged propositional attitude concepts like "belief" which are inherently aspectual: Ascription of a belief about a given object to an agent is sensitive to the aspects under which the agent subjectively represents the object in question. To ascribe the belief "Clark Kent is at home" to an agent is fundamentally different from ascribing to her the belief "Superman is at home".

Empirically, these differences in the representational repertoire of the two systems should thus manifest themselves in distinctive and differential patterns of performance. S1 should have characteristic signature limits such that on the basis of this system, agents can master (only) those FB tasks that can be solved by tracking agents' purely relational attitudes. This will apply to tasks for which it is not strictly required to grasp the aspectuality of mental representation. Level-I-perspective-taking tasks fall into this class, for example. In such tasks one merely has to track *whether* someone has seen an object or not, but not how she has seen that object. Similarly, many simpler change-of-location FB tasks fall into this class, too. Here, subjects only have to keep track of *which* events a protagonist has or has not registered (and not *how* she has represented these events). In a classical change-of-location FB task (Wimmer & Perner, 1983), the protagonist at time 1 puts an object into box 1, which is then at time 2 transferred in her absence to box 2, and the crucial question is where the agent, upon return at time 3, will search for her object. In order to solve this task, an infant may only need to represent that the agent at time 1 registers (stands in perceptual contact with) O in box 1. Since the protagonist subsequently does not register any other or competing information, this registration is not updated, and thus the infant can predict at time 3 that the agent will act on the basis of this registration. But since registration is a relational attitude, it does not allow the infant to distinguish how the protagonist may have represented the object and thus would not allow mastery of tasks that would require such a more fine-grained understanding.

S2, in contrast, should not be subject to such signature limitations. Rather, it should enable the mastery of a great variety of tasks the common denominator of which is that they require an understanding of mental representation and its aspectuality. That is, subjects operating with S2 should be able to solve standard change-of-location FB tasks just as much as more complex tasks that require explicit representation of aspectuality (such as answering the question "Where does Peter believe Clark Kent is now?"). While there should thus be divergence and dis-unity in different types of implicit FB tasks (such that young children consistently master those FB tasks that do not strictly require an understanding of aspectuality but fail those that do), for explicit FB tasks there should be convergence and unity (such that all kinds of such tasks begin to be mastered at the same time and in correlated fashion).

1.3. The empirical situation so far

Turning first to young children's performance in implicit FB tasks of various types and topics, is there any evidence for disunity and dissociation? The empirical situation so far is complex. On the one hand, some studies suggest that infants and toddlers are able to solve some implicit FB tasks that require an understanding of aspectuality around the same time that they master implicit non-aspectual change-of-location FB tasks (Buttelmann, Suhrke, & Buttelmann, 2015; Scott & Baillargeon, 2009; Scott, Richman, & Baillargeon, 2015). For example, in Buttelmann et al. (2015), a protagonist reached toward an object with misleading appearances (e.g. an A that looked like a B). In some cases she was aware of the real nature of the object and thus knew that it was an A that only looked like a B (TB condition) whereas in other cases she was not aware of the true nature of the object and thus took it by its appearance as a B (FB condition). Infants then, in some cases, and for some sub-sample of the stimuli, helped the protagonist differentially in TB and FB conditions (they tended to give her another B-object in the TB condition more often than in the FB condition, and tended to give her another A-object in the FB condition more often than in the TB condition).

These studies taken by themselves, however, are very difficult to interpret. One reason is that all of them have used a single isolated vignette each of which leaves room for alternative, more parsimonious explanations, either in low-level terms (Heyes, 2014a, 2014b) or in terms of children's tracking belief-like states rather than fully-fledged aspectual beliefs (e.g., Butterfill & Apperly, 2013). In the absence of