British Journal of Developmental Psychology (2012), 30, 59–74

The British Psychological Society

www.wileyonlinelibrary.com

Do infants have a theory of mind?

Hannes Rakoczy*

Institute of Psychology & Courant Research Centre 'Evolution of Social Behaviour'
University of Göttingen, Germany

The central question debated in current research on infant social cognition is 'do infants have a theory of mind?' It is argued here that this question is understood and treated in radically different ways by different participants of the debate arguing either for (e.g., Onishi & Baillargeon, 2005) or against early competence in theory of mind (e.g., Perner & Ruffman, 2005). As a consequence, there is considerable talking past each other, both sides make claims that appear incompatible but are actually answers to different questions and framed at different levels of description. Some conceptual distinctions from the philosophy of mind are therefore introduced to describe the different interpretations of the question and the misunderstandings based thereupon, with the aim of providing some conceptual clarification as groundwork for future debates.

Do infants have a theory of mind (ToM)? This is one of the most hotly debated questions in recent developmental research against the background of findings that children in implicit tasks reveal some sensitivity to belief situations in infancy but only solve explicit tasks about belief understanding around 4 years (for review, see Baillargeon, Scott, & He, 2010). The aim of the present paper is not to answer that question. What I would like to suggest, rather, is that the question is somewhat ambiguous and that in the debate about it there is considerable talking past each other: researchers pro and con infant theory of mind respond actually to different readings of the question, giving answers seemingly in conflict with each other but actually quite compatible – answers to different questions.

To avoid such misunderstanding, more conceptual clarity is needed. In this paper, therefore, I will review some central conceptual distinctions from recent philosophy of mind – distinctions between different types of representation. These distinctions will be applied to infant ToM at two levels: first regarding the type of attitude infants take towards others' mental life (do they believe something about others' mental states, or do they engage in simpler kinds of states?), and second, regarding the kind of attitude they ascribe to others (beliefs, or some simpler kinds of state?).

*Correspondence should be addressed to Hannes Rakoczy, Institute of Psychology, University of Göttingen, Waldweg 26, D-37073 Göttingen, Germany (e-mail: hannes.rakoczy@psych.uni-goettingen.de).

DOI:10.1111/j.2044-835X.2011.02061.x

Some distinctions between levels of description and kinds of representation

When wondering whether an individual, in particular a pre- or non-verbal creature has certain cognitive capacities, it is easy to slip into talk that is inherently ambiguous regarding the level of description and the type of cognitive phenomenon referred to. 'Do children *know* the rules of their native language?' or 'Do infants *understand* that objects persist over time?' are natural questions, much asked in developmental psychology, but ones in much need of conceptual clarification: what *knowing* and *understandin*g here is supposed to mean is by no means clear and unambiguous. The most fundamental reason for this is probably that on the one hand the terms are taken from our folk psychology where we all understand them perfectly. On the other hand, the terms often have technical counterparts in cognitive science that sometimes start off from the original folk psychological meaning but then deviate considerably from it. Philosophers of cognitive science have therefore suggested a number of (partly overlapping and similar) conceptual distinctions to avoid such confusions.

Personal versus subpersonal levels of description

One of the most fundamental distinctions is between different levels of description and explanation (Dennett, 1969). When we apply our everyday folk psychology, we describe each other as persons, as beings capable of mental states – in particular, propositional attitudes with semantic content (such as beliefs, desires, and hopes). Crucially, we ascribe these states to the whole person. And we explain actions of the whole person by reconstructing the reasons (usually the beliefs and desires) that rationalize the actions. These distinctive modes of description (in terms of propositional attitudes and the like) and explanation (rational explanation) at the personal level contrast with the subpersonal level. At this level we descend to information-processing and neurophysiological descriptions and explanations common in cognitive science, and describe parts of the information processing system or the brain. While often there is no confusion of these levels with the personal level because special technical vocabulary is used, some of the terms used in subpersonal descriptions have their original home in our folk psychology and thus create some ambiguities (a prominent example is 'representation' – something people have in our folk psychology, but that information-processing subsystems store and manipulate according to subpersonal theories).

Propositional attitudes versus subdoxastic states

Related to the distinction between levels of description, there is a distinction between different kinds of cognitive states: the propositional attitudes of our folk psychology, above all beliefs, are distinguished from the so-called 'subdoxastic states' of computational scientific theories (Davies, 1989; Stich, 1978). We *know* the rules of chess, and we are said to *know* abstract grammatical rules applying to the deep structure of sentences. Does 'know' have the same meaning in these two cases? It seems that it does actually not, that we are dealing with normal propositional attitudes, in particular (justified true) belief, in the former case, but with some informational state falling short of being a belief in the latter case – so-called 'subdoxastic' (doxastic = pertaining to belief) states. Subdoxastic states do play a role in the causal history of regular beliefs. For example,

the subdoxastic states pertaining to grammatical deep structures clearly play a role in the production of the regular belief that a sentence is ungrammatical. But they are not themselves beliefs.

Regular beliefs differ from subdoxastic states in at least three respects (Davies, 1989; Stich, 1978):

(a) *Inferential integration*: Beliefs are inferentially integrated, or – as it is sometimes put – inferentially promiscuous: beliefs (e.g., that 'p') combine with other beliefs (e.g., that 'if p then q') to inferentially yield – in theoretical reasoning – further beliefs (e.g., that 'q'). And beliefs combine with other attitudes in practical reasoning to yield decisions and actions. Subdoxastic states, in contrast, while they do interact in some limited ways, fail to exhibit this inferential integration. Subdoxastic informational states about grammatical rules, for example, do combine with other such states in the process of language comprehension and production. But they are not inferentially integrated with informational states pertaining to areas other than grammar, and in particular they are not integrated with regular beliefs.

(b) *Accessibility to consciousness*: Beliefs normally are accessible to consciousness. When we believe that *p*, we can usually access the content *p* and use it, for example, in asserting 'p' and asserting 'I believe that p'. We are not conscious, in contrast, of the content of deep grammatical rules, and we therefore cannot use these contents in making assertions or self-ascriptions in the way we can with beliefs.

(c) *Conceptualization*: Beliefs necessarily have conceptual content. Subdoxastic states, however, often have informational content that falls short of being properly conceptual. The information-processing system dealing with deep grammatical structure, for example, might track such things as *Wh-movements* (syntactic operations involved in forming questions) or the like, and the grammarian does have a concept of 'Wh-movements'. But of course that does not mean that the informational content of the states tracking Wh-movement in normal speakers needs to be conceptual content.

These last two respects (access consciousness and conceptualization) are important distinctions in themselves and should therefore be dealt with in more detail.

Conceptual versus non-conceptual content

Propositional attitudes have conceptual content. We ascribe beliefs, desires, hopes, and wishes that some object *a* have some property *F* (e.g., 'Peter is tall'). The content of such attitudes is thus 'a has F' and having such attitudes requires the possession of the concepts 'a' and 'F'. Many information-processing states postulated in cognitive science, in contrast, have informational content that is best described as non-conceptual. In language perception, for example, we undergo informational states tracking things like voice onset time. In some sense these states are about voice onset times, have information about such matters as their content. But this is not content involving the concept 'voice onset time'.

Why not? What is missing? What concepts are, is of course a notoriously difficult question much debated in philosophy and psychology, one that we certainly won't be

able to answer here. But even in the absence of a final definition of 'concept' there are good basic criteria for distinguishing conceptual from non-conceptual content.

(1) Concepts are structured abilities that freely recombine, and such recombination is *compositional*: the meaning of thoughts is determined by the meaning of the concepts of which the thought is made up and the way they are put together.
(2) Such recombination is *productive*: out of a finite number of concepts potentially infinitely many thoughts can be composed.
(3) Relatedly, concepts meet what has been dubbed the *Generality Constraint* (Evans, 1982): if a thinker can think 'a has F', and 'b has G', she must *ipso facto* be able to think 'a has G', 'b has F', etc.

Compositionality, productivity, and generality, however, do not necessarily apply to non-conceptual content. For example, there might well be informational states that are sensitive to fine-grained differences in the duration of events differing in voice onset time – states that thus are in some sense about duration. But such states might well exist without the informational system being able to track the duration of anything else than events differing in voice onset time. Such isolated capacities, however, are ruled out on the level of conceptual content: there is no such thing as a thinker who can in principle apply her concept 'duration' to one kind of event only.

(Access-) conscious versus non-conscious states

Consciousness, many have argued, is a cluster concept actually used to make several partly independent distinctions (e.g., Block, 1995). The most relevant distinction for present purposes is between states that are or are not *access-conscious*. The propositional attitudes we ascribe in our folk psychology, typically are access conscious[1]. Roughly, 'a state is access-conscious if, in virtue of one's having the state, a representation of its content is (1) inferentially promiscuous [. . .] i.e. poised to be used as a premise in reasoning, and (2) poised for (rational) control of action and (3) poised for rational control of speech' (Block, 1995, p. 231). For example, access-conscious occurrent beliefs that *p* are beliefs that are inferentially promiscuous (I can freely use *p* as a premise in reasoning; criterion 1), that control my practical reasoning (criterion 2), and that I can use in rational speech, for example, in asserting that *p* (criterion 3).

Clearly, many of the information-processing states postulated by cognitive science at the subpersonal level are not access conscious: I cannot use the (non-conceptual) content 'this sound has such and such voice onset time' as a premise in any kind of reasoning and I cannot use it rationally in speech.

Applying the distinctions to infant theory of mind

Let us now apply these distinctions to the question under discussion 'Do infants have a theory of mind?' I will mainly concentrate on the distinction between propositional attitudes proper (above all beliefs) and subdoxastic states for two reasons: *first*, the belief-subdoxastic distinction is a very comprehensive distinction (incorporating

[1] *At least in their occurrent form: occurrent attitudes (in contrast to dispositional ones) are attitudes we momentarily and actively hold, for example, in making a judgment (e.g., that there is vanilla chocolate ice-cream in the freezer) or in entertaining a desire in a choice situation (e.g., that the vanilla ice-cream should be the dessert).*

inferential integration, conceptualization, and potential access consciousness)[2]. *Second*, the belief-subdoxastic distinction is particularly helpful for dealing with the question of intentionality of second order or meta representation on the part of the infant (believing about believing) as it applies at following two levels.

(1) In observing others, what does the child herself do: does she have a belief about others, or just some belief-like state? (the first-order question).
(2) And what does the child *ascribe* to others: beliefs proper, or just some belief-like state? (the second-order question).

To distinguish these questions at the different levels, it will be helpful to consider the problem in its general schematic form. To formulate the matter as neutral as possible and without prejudging the issue how sophisticated infant ToM really is, what we are dealing with is an observer with some cognitive relation (R_1) to the situation that some protagonist has some cognitive relation R_2 to a situation (p).

Observer R_1 [protagonist R_2 (p)]

In proper intentionality of second order, R_1 and R_2 are normal propositional attitudes, like in 'Eve suspects that [Adam believes (it was all Eve' fault)]'. But what are R_1 and R_2 now in the case of infants?

What the infant does in observing others: What is R_1?

What kind of state or attitude is R_1? Is it a personal-level state ascribable to the whole infant? Or are we dealing with something simpler, perhaps some information processing state to be ascribed to some part of the infant's cognitive machinery at the subpersonal level?

It is here, I think, that much of the debate involves some talking past each other due to different terminological commitments: some say 'Yes, the infant has a ToM' and thereby mean to refer to subdoxastic states (perhaps of a ToM module), whereas others deny that infants have a ToM, understood as beliefs about beliefs – but of course these are two perfectly compatible claims.

Beliefs, remember, differ from subdoxastic states in at least three respects: (1) beliefs usually are (at least potentially) access conscious, (2) they are inferentially integrated and promiscuous, and (3) they involve conceptual content. Let us deal with (1) and (2) first, are infants' ToM capacities access conscious (even potentially) and inferentially integrated?

---

[2] *A word should be said at least why I'm not framing the debate about infant ToM with the use of some implicit-explicit distinction. The reason is that there are too many such distinctions, quite diverse ones (e.g., implicit beliefs as implied beliefs, implicit memory as unconscious recollection, implicit knowledge as procedural know-how, etc.) without any one being the agreed upon distinction. Therefore, it is of little explanatory help, for example, to just call some precocious abilities in infants 'implicit XYZ' (implicit ToM…) without further qualification. More specific contrasts, however, can be explanatorily very helpful, but it is an open question whether they are so because they rest on more fundamental distinctions like the ones between (access-) conscious versus non-conscious, declarative versus procedural, or between different kinds and formats of representations and their inferential integration (see Dienes & Perner, 1999).*

Empirical issues

What the data so far suggest is that infants clearly, to put it as neutrally as possible, are sensitive to certain belief-involving situations. They distinguish some situations in which a protagonist has a false belief from situations in which she has a true belief. The main evidence for this comes from their looking behaviour, both in violation of expectation studies (e.g., Onishi & Baillargeon, 2005; Surian *et al.*, 2007; Träuble, Marinovic, & Pauen, 2010) and in anticipatory looking (Southgate, Senju, & Csibra, 2007).

Such sensitivity to belief situations, however, does not imply beliefs about beliefs. The main reason is that mere looking behaviour by itself seems too isolated a behaviour, and too non-voluntary for that, to reveal an access-conscious, inferentially promiscuous state. Access-conscious and inferentially promiscuous states reveal themselves in diverse contexts, in different theoretical and practical inferences, in the rational control of action and speech. Given the empirical evidence on standard ToM tasks and other related tests, there is not much controversy about the question whether 4-year olds have beliefs about beliefs: the reason is that they reveal ToM capacities in flexible ways in different contexts and different kinds of theoretical and practical inferences, and in their rational control of action and speech. These concerns, of course, reflect a more general concern regarding the interpretation of looking time data by themselves: how is a precocious cognitive capacity to be described and interpreted that reveals itself only in one very limited context before it appears – often years later – in more action-based and verbal tasks?

From an empirical point of view, what would be needed to tell whether infants' sensitivity to belief-involving situations amounts to beliefs about beliefs or not, are two things: *first*, more data with measures of sensitivity to belief-involving situations other than looking time; and *second*, good theoretical models of what might count as appropriate indicators of and convincing evidence for inferential promiscuity and access consciousness on a preverbal level, given that our main criteria for them rely heavily on language (e.g., Bermúdez, 2003): from a practical point of view, the clearest evidence of access consciousness of a state is the person's ability to use the state in the rational control of speech (i.e., to express and self-ascribe the beliefs in question) and the best evidence of inferential promiscuity is the ability to reason verbally in promiscuous ways (Block 1995).

First preliminary evidence for sensitivity to belief-involving situations with measures other than looking time have recently been found in two studies making use of children's communication and helping behaviour (Buttelmann, Carpenter, & Tomasello, 2009; Southgate, Chevallier, & Csibra, 2010). One of the big open empirical questions for future research is thus to what degree such results generalize. How context general and flexible is early belief sensitivity, and in particular is it so to a degree that merits an interpretation as access-conscious and inferentially promiscuous beliefs about beliefs?

Theoretical issues

From a theoretical point of view, the personal/subpersonal and belief/subdoxastic distinctions can be helpful in clarifying some fundamental issues of debate. One is the question whether infants' failure to solve explicit ToM tasks reflects a competence failure or merely a performance problem. Proponents of infant ToM usually argue for some competence-performance distinction (e.g., Leslie, 2005): infants already have the cognitive competence (representing beliefs) but suffer extraneous performance problems in explicit tasks. Leslie and colleagues, for example, have supplied sophisticated information-processing models distinguishing between core ToM representations, supposed to be intact early, and extraneous factors such as executive function (EF)

developing in more protracted fashion (e.g., Leslie *et al.*, 2004). While these models are among the best-formulated information-processing theories of ToM we have, there is one potential conceptual unclarity obscuring the debate.

It makes good sense in subdoxastic accounts to distinguish between core representations (e.g., Leslie's ToMM representations) and extraneous processes (e.g., EF) in putting these representations to work. But such subdoxastic distinctions do not necessarily map 1:1 to personal-level descriptions. To assume this would be to commit a 'mereological fallacy' (Bennett & Hacker, 2003), confusing properties of parts of the information processing system (e.g., tokening a ToMM representation) with personal-level properties of the whole system (e.g., believing that someone believes that *p*). And it is essential to personal-level states, in contrast to subdoxastic ones, that they do their work in reasoning and the rational control of action. In other words, if a subpersonal state doesn't do any job in the performance of rational action, etc., then it simply is not the subpersonal realizer of a personal-level belief. The state's not doing any work in the rational control of action, etc., is not a mere performance problem. Rather, 'performance limitations' so severe turn it into a very competence problem.

This becomes particularly clear on functionalist analyses according to which mental states are what they are due to their functional role in our mental economy (e.g., Putnam, 1960; Sellars, 1956). Propositional attitudes in particular are essentially defined, *inter alia*, through their role in theoretical and practical reasoning and the rational control of action and language. They are thus abstractly functionally characterized states multiply realizable in different ways in different creatures (e.g., neurally in us). A crucial distinction regarding realization is between *core realizers* and *total realizers* (Shoemaker, 1981). Let's assume, taking a philosophers' toy example, that pain is realized in normal humans in C-fibre firing (which is to say, that the C-fibres are so hooked up that the functional role characteristic of pain is fulfilled) Then each token C-fibre firing in a normal human realizes a token pain experience. But this is only the case in the context of a normally functioning organism. Take out the C-fibres and let them fire in a test tube and there will be no pain. C-fibre firing is thus a *core realizer* of pain, but only C-fibre firing plus its essential connections to other parts of the system is the *total realizer* of pain.

Applied to information-processing accounts of ToM, what we should say is the following. In normal adults some ToMM representation might well be the *core realizers* of beliefs about beliefs because they are functionally related to other subdoxastic states in the right ways (playing their role in theoretical and practical reasoning and the rational control of action, etc.) so that the whole system is a *total realizer* of beliefs about beliefs. But that simply does not mean that these representations already realize beliefs about beliefs in infants. Rather, the infants might have the subdoxastic parts, but not the right kind of functional connections yet, that constitute the very competence in question (compare the C-fibre in the test tube).

We should thus be careful to distinguish between the claim that infants/adults have ToMM representations (a claim at the subpersonal level), and the personal level claim that infants have beliefs about beliefs. The former might well be true with the latter being false. All of this in no way speaks against such information-processing accounts as Leslie's, of course. On the contrary, it helps clarify their true explanatory merits: they supply subpersonal realization explanations (Cummins, 1983) of the personal-level capacity to have beliefs about beliefs.

To summarize, recent studies suggest that infants are sensitive to belief-involving situations. But it is a further question whether such sensitivity is best characterized as involving a (personal level) second-order belief, finding only limited expression due

to performance problems. Alternatively, the sensitivity might be better described as involving simpler subdoxastic states still failing to fulfil some essential properties of beliefs proper (particularly, inferential integration and access consciousness). More systematic data and better theoretical accounts of non-verbal analogues of standard criteria for inferential integration and access are needed to settle this question.

What the infant ascribes to others: What is R_2?

From the nature of R_1, let us turn now to the nature of the supposed content of this state, namely [protagonist R_2 (p)]. This is what most of the debate in this area has been concerned with, for example, in disputes about whether the infant is really sensitive to others' beliefs, or just to simpler informational states or even just to behaviour (see, e.g., Apperly & Butterfill, 2009; Perner, 2010; Perner & Ruffman, 2005; Sirois & Jackson, 2007).

The Generality Constraint

Propositional attitudes, beliefs in particular, have conceptual content that is compositional and productive and conforms to the Generality Constraint: if one can apply a predicate F to an object a, one must be able to think in principle that other objects b and c have or do not have F, etc. The Generality Constraint applies in specific ways to mental state predicates (Evans, 1982; Strawson, 1959): if I can ascribe a kind of state, say a belief with a certain content p to some person X, then I must be able in principle to ascribe (1) the same kind of state to other persons Y and Z and (2) the same kind of state with different contents q and r. In our schema, if R_2 is supposed to be a proper mental state predicate, 'belief' in particular, then there must be some generality regarding the 'protagonist' and the 'p' slots.

Generality in ascribing thought contents. If one is capable of ascribing beliefs, one must in principle be capable of doing so with a variety of contents – in fact, theoretically with all the thought contents one can entertain oneself. If I can think 'there are no daffodils on Mars', for example, and have the concept of 'belief', I am *ipso facto* in a position to think, 'I think (there are no daffodils on Mars)'. Now, most of the initial findings with infants concerned the supposed ascription by infants to others of beliefs about object locations (e.g., Onishi & Baillargeon, 2005; Surian *et al.*, 2007). If the findings remained limited to such narrow contexts, it would be very difficult or even impossible to tell whether the content aspect of the Generality Constraint was fulfilled. More recent findings, however, suggest that infants are also sensitive to belief-involving situations regarding contents other than just object locations such as object properties (e.g., Scott & Baillargeon, 2009). Considering the performance of 4-year olds on explicit tasks, children show comparable competence in tasks with a wide variety of topics, such competence being very robust relative to superficial task variations (e.g., Wellman, Cross, & Watson, 2001). In fact, this is one central reason why many researchers are confident that a truly general conceptual competence is measured by such explicit tasks and why a real developmental shift is tapped around 4 years. It will thus be a crucial question for future infancy research whether more evidence for a similar generality regarding ascribed thought contents can be found in more implicit tasks already.

Generality in ascribing thoughts to protagonists. The other aspect of the Generality Constraint applying to mental state terms concerns the protagonist: if one is capable of ascribing beliefs, one must in principle be capable of doing so to a variety of protagonists. In particular, one must be capable of doing so to third persons and oneself alike. One interesting aspect about this is that propositional attitude predicates (e.g., 'belief') have unitary meaning when applied in the first and in the third person, but the ascription itself works in radically different ways in these two cases: we ascribe such states to others on the basis of (behavioural) evidence, but we ascribe them to ourselves (at least in the case of present occurrent states) without such reliance on behavioural criteria (Evans, 1982; Shoemaker, 1968)[3].

Historically, the main concern has been with 'self-knowledge first' accounts of mental state ascription (claiming that we first ascribe mental states only to ourselves, and then – by reasoning from analogy to our own case – ascribe them to others derivatively). The Generality Constraint and precursor ideas have been used to argue that such accounts are incoherent. There is no such thing as a predicate only applicable in principle to one object, that is, oneself (Strawson, 1959). The present case – the question whether infants fulfil this aspect of the Generality Constraint – might be an interesting case in the other direction. All of the evidence so far on infants' sensitivity to belief-involving situations concerns third persons. This is in contrast to classical explicit ToM tasks where children begin to master mental state ascriptions to others and themselves in tandem (e.g., Gopnik, 1993; Perner, Leekam, & Wimmer, 1987; Rakoczy, 2010). This might actually be another reason that there is not much debate about whether 4-year olds really have acquired a proper concept of 'belief'. Not only do they show general and flexible capacities in different kinds of tasks of ascribing diverse belief contents to protagonists; they also show flexibility and generality in ascribing beliefs to others and themselves alike.

It is of the essence of propositional attitude concepts that we can apply them to both first and third person, and that we do so in the first person in characteristic, non-inferential ways (e.g., Shoemaker, 1968). It is thus one of the empirical challenges for infancy research to find out whether there is any evidence that infants can apply their hypothesized 'belief' concepts to themselves in the ways characteristic of self-ascription, analogous to such findings on explicit tasks with 4-year olds. If infants were in principle limited in their ascription of propositional attitudes to third persons, skepticism about whether what they are applying are *mental state* concepts at all would be justified[4].

---

[3] *This last point does not, as is sometimes mistakenly assumed, imply any dubious Cartesian picture of the mind and introspection. In fact, it does not require any special faculty of introspection conceived of as inward perception at all. Rather, the non-inferential self-ascription of occurrent mental states rests on an intimate connection between the expression of first-order thoughts and the second-order self-ascription of such first-order thoughts: when self-ascribing the thought that p, what I am doing is neither looking inward nor making inferences from my behaviour. Rather, I look into the world, so to speak, wonder whether p is the case, and by reaching a first-order opinion on the matter, expressible as 'p', I can ipso fact self-ascribe 'I believe that p' (e.g., Shoemaker, 1995). In various places this procedure has been dubbed 'Evans' Procedure' after a much quoted passage from Gareth Evans: 'I get myself in a position to answer the question whether I believe that p by putting into operation whatever procedure I have for answering the question whether p' (1982, p.225). This intimate connection between occurrent first-order thoughts and their self-ascription reveals itself also very clearly in the so-called 'Moore's Paradox': sentences and thoughts of the form 'p, but I believe that non-p' are paradoxical.*

[4] *If the infant were so limited, she would be in the following situation: she would have beliefs, among them occurrent ones. And she would have the concept 'belief'. But she would be unable in principle to apply this concept to herself. She would be self-blind, so to speak. When believing that p and when asked about her beliefs whether p is the case, she would thus in principle have to think 'p, but no idea whether I believe that p'. Such a scenario has been taken to be incoherent by many philosophers (e.g., Shoemaker, 1995) – showing that the premise that the infant has the concept 'belief' must be false.*

Specific requirements of ascribing beliefs

Besides formal requirements for mental state ascriptions generally stemming from the Generality Constraint, there are more particular requirements for the ascription of beliefs specifically. Is R_2 a belief, or might it be some similar but simpler state?

Basically, similar considerations apply here at the second order (are what the infant ascribes to others beliefs?) as the ones we already discussed at the first order (is the attitude the infant takes towards others best considered a belief?). Beliefs are what they are due to their functional role, the role they play in a mental economy: their relation to input (normally, perceptual encounters with objects in standard conditions lead to true perceptual beliefs, etc.), their relation to other mental states (beliefs lead to other beliefs via theoretical reasoning, and together with desires lead to decisions and intentions in practical reasoning, etc.), and their relation to output (beliefs in practical reasoning lead to decisions which usually lead to action, etc.). By their nature, beliefs are attitudes with propositional content, that is, they are semantically evaluable, and they are subject to certain normative constraints and considerations, in particular norms of truth (beliefs aim at truth) and consistency (beliefs rationally ought to be consistent with each other) (e.g., Velleman, 2000). Describing another person's state as a beliefs involves, in a famous phrase, 'placing it in the logical space of reasons' (Sellars, 1956, p. 169), a space essentially defined by standards of rationality and truth.

Empirical issues. The central empirical challenge for infant ToM research is thus:, is there any good evidence that infants are capable of tracking the functional roles of beliefs and their semantic properties and normative constraints? Or are infants sensitive only to simpler states that share part of the functional role of beliefs?

This is an old question, of course, in ToM research, with many sophisticated analyses of how children develop from understanding simpler states to understanding beliefs proper and much corresponding empirical evidence (e.g., Perner, 1991). Most of the analyses centre around distinguishing simpler types of attitudes from beliefs where the former share some functional features with the latter but are fundamentally different in that they lack the possibility of misrepresentation. Young children before classical ToM age have been characterized, for example, as confined to level I (in contrast to level II) perspective taking.

Whether a subject operates with the concept of 'belief' proper or with some such notion of simpler states that might have considerable overlap regarding their functional roles with beliefs is a difficult question. It can only be decided on the basis of much convergent evidence (such that the subject is systematically sensitive to belief-involving situations proper, and not just to situations that can be described as involving beliefs *or* as involving such simpler states alike). Such converging evidence in the case of 4-year olds makes most researchers in the field confident that children at this age really do track beliefs and not just belief-like states.

Given that the data on infants are still quite sparse, no such consensus of confidence has been reached yet regarding infant ToM, however. Apperly and Butterfill (2009), for example, have recently proposed a two-systems account of ToM, with an early developing System 1 limited to the ascription of belief-like states, and a later developing fully fledged System 2 for ascribing beliefs and other propositional attitudes. They argue that existing data do not allow us yet to distinguish whether infants operate with a concept of belief or with a concept of belief-like states such as 'registration'. The latter is understood as an informational state sharing many of the functional aspects of beliefs,

even allowing for simple forms of misrepresentation, but lacking the semantic and normative richness, the flexible role in reasoning, etc., characteristic of beliefs.

Apperly and Butterfill describe the situation as analogous to early numerical development, prima facie, when considering infants' dealing with small sets, it looks like they had a concept of number and were capable of simple numerical operations such as addition and subtraction (e.g., Wynn, 1992). But on closer inspection these remarkable capacities to track sets and perform simple numerical operations turn out to be due to a much simpler object individuation system that works for sets that can be individuated in parallel (<4) and thus has a very clear signature limit (see Carey, 2009).

Regarding infant ToM, more data are needed to explore whether analogous signature limits characterize infants social cognition such that they ascribe merely belief-like states that are not specific and flexible enough in their functional roles in reasoning, etc., to count as beliefs. To take just one example, the most impressive finding in terms of age of the participants to date comes from a study with 7-month-olds. In this study, infants (and adults) showed automatic processes of representing a protagonist's representations towards reality. Infants were influenced in an object recognition task by what a protagonist had previously seen being the case, looking longer in cases of a mismatch of the content of this representation and reality as represented by the infant herself (Kovács, Téglás, & Endress, 2010). This finding has been taken by some (e.g., Gergely, 2011) to show that infants represent the protagonist's belief. But while this paradigm clearly shows differential response as a function of some representational content on the part of the protagonist, it does by no means show that infants represent the protagonist's belief qua belief. Why not? Because there simply is no indication that they ascribe to the protagonist any mental state with a specific functional role remotely similar to that of beliefs. We do not know, for example, whether the same pattern of response time interference would also be found when the infant had seen someone with a desire, hope, or some other propositional attitude with a content deviating from the real state of affairs. In other words, infants somehow associate some content with the protagonist, but whether they ascribe this content as the content of an attitude as specific as belief, we cannot tell from such data alone. More evidence would be needed.

Finally, what about children's understanding of the normative dimensions of beliefs? This is one of the aspects that have received relatively little attention, even in classical ToM research (see, e.g., Koenig, 2002). But we do have some indications that toddlers have some grasp of the normative dimension of cognitive states and speech acts, of their different normative 'directions of fit' (Searle, 1983): beliefs aim at truth (mind to world direction of fit), desires aim at fulfilment (world to mind direction of fit). For example, when confronted with speech acts with different directions of fit (assertions versus imperatives) with the same propositional content, they respond very differently in cases of the non-fulfilment of the semantic content of the speech act. They criticize the speaker for being wrong in the case of assertions, but criticize the addressee for making action mistakes in the case of unfulfilled imperatives (Rakoczy & Tomasello, 2009).

A proper concept of belief requires some awareness of such normative issues. Finding out whether infants have some such awareness will be another big challenge for future research.

Theoretical issues. Even if some of these empirical questions were answered, however, some rather fundamental theoretical issues would remain. Let us suppose we knew that infants are systematically sensitive to many situations involving states with the

functional profiles and the semantic and normative aspects of beliefs: they distinguish such situations from situations that do not involve states with the functional profile of beliefs. What does this show? In particular, does this show that infants have the concept 'belief?' This is a tricky question, one that depends on a whole lot of background assumptions about what concepts are. And it is here, I suggest as another diagnosis, that much talking past each other happens due to different concepts of 'concepts' in play.

The main dividing line here is between atomistic accounts that view concepts as unstructured and atomistic, much like pointers towards the world, on the one hand, and functionalist and holistic accounts on the other hand that view concepts as structured and defined by the role they play in a network of other concepts with which they are semantically and inferentially related (so-called conceptual role semantics).

Concept atomism

Some proponents of the claim that infants operate with a concept of belief, Alan Leslie in particular, argue for such a claim against the background of an atomistic account of concepts (e.g., Leslie, 2000). Concepts generally, on such accounts are viewed on the model of natural kind concepts. Quite plausibly, natural kind concepts (such as 'oak', or 'tiger') do not carry much descriptive meaning for most of us – we cannot say much about oaks or gold and cannot give any defining features *a priori*. What we can do, however, is refer to oaks and gold, pick out such things in the real world, usually based on characteristic surface features (sometimes called 'stereotypes') correlated with and caused by the essential hidden features of the kind in question. And we can do so, according to the causal theory of reference, because there is a causal connection between the natural kind and our use of the corresponding term or concept (Kripke, 1972; Putnam, 1975). That is, we can pick out and refer to 'oaks', say, without being able to make any sensible conceptual inferential connections between 'oak' and anything else. We can find out the essential properties of oaks only by *a posteriori* scientific study, not by conceptual analysis because 'oak' simply does not have any essential conceptual connections to other concepts.

While such a semantic theory is considered plausible by many for natural kind concepts, the view that other kinds of concepts function like this is much more contentious. Leslie, for example, makes such assumptions explicitly for mental state concepts (e.g., Leslie, 2000). Our concept of 'belief', or 'desire' say, work much like our concept of 'oak': we are able to track beliefs or desires, but we might be able to do so initially without having any conceptual connections between 'belief', 'desires' and any other concepts. In fact, this is how, on this picture, the infant is to be portrayed. In development we then come to acquire, according to Leslie, such connections by finding out things about beliefs and the like (much like we find out about oaks and gold), but this is a secondary process that leaves the concepts themselves unchanged.

Such a position has some obvious problems. To name just two: *first*, it seems to run counter to our conceptual practices. We would not know what to make of a person who was supposed to use the concept 'belief' to pick out belief-involving situations, but were unable in principle to make any conceptual and inferential connections involving the concept 'belief' (e.g., to 'truth'). Such connections clearly seem essential, constitutive to our very mental state concepts. *Second*, it is unclear how a causal theory of reference is supposed to work in the case of mental state concepts. In the case of natural kind concepts, the basic story of causal theories of reference is roughly the following. There are deep essential features of natural kinds that are causally responsible

for some superficial appearance features (stereotypes) associated with the natural kind. Under normal circumstances these superficial features can be picked up by our sense organs, and we learn to apply the natural kind concept to instances with these features discriminatively and reliably. In this way there is a reliable causal connection between the kind and our use of the natural kind concept. But how should such a story work in the case of mental states such as beliefs? There are no specific superficial features associated with the class of belief-involving situations. Beliefs just are too abstract to go along with a specific appearance.

Moderate concept holism

Moderately holistic accounts of concepts, in contrast, see concepts as, at least partly, constituted by their semantic and inferential relations in a network of interdependent concepts (e.g., Block, 1986; Sellars, 1963). In developmental psychology, the theory is a prominent kind of such an account (e.g., Gopnik & Meltzoff, 1997). Applied to mental state concept specifically, such accounts claim that our concept of 'belief' is what it is due to its positions in a web of concepts ('belief', 'desire', 'action', 'rationality', 'truth'. etc.) semantically and inferentially related. In fact, according to this theory, the very same constraints operative at the first order (which functional roles make a state one of belief?) apply now at the second order (which functional roles of a state must an ascriber represent for her to possess the concept 'belief?'). When, on such a reading, one credits a child with the concept 'belief', one thereby claims that the child has mastered essential conceptual connections between 'belief' and other concepts constitutive of our folk psychological notion of such states.

Regarding cognitive development, such theories have the merit that they allow us to describe gradual conceptual development and conceptual change, by describing different stages in which different types of inferential connections, for example, involving the notion of truth and (mis-) representation are mastered (see, e.g., Perner, 1991).

Some severe problems, however, plague such holistic accounts (e.g., Fodor, 1998): if holism is taken in its radical form (a concept is constituted by *all* of its inferential connections), absurd consequences follow. No two people can ever share a concept (they will never have exactly the same inferential dispositions in place in total), and no person can be said to have the same concept over time (as she constantly changes her inferential dispositions, e.g., by learning new facts). If holism is taken in more moderate forms (a concept is constituted only by some of its inferential connections), a criterion is needed of what makes some inferential connections constitutive and others merely accidental.

Leaving aside the general pros and cons of these different concepts of concepts (it is not the aim here to argue for one or the other, anyway), what should be noted is how vastly different conclusions to the question "Do infants have a concept of 'belief'?" they yield. An infant reliably tracking some belief-involving situations without any inferential liaisons between 'belief' and other notions ('truth', etc.) would count as having the concept 'belief' according to Leslie. But according to holistic accounts of concepts, such an infant would either not count as having any kind of conceptual (rather than perceptual) capacity at all (no inference, no concept!), or at most as having a concept of some state much simpler than belief. Considerable parts of the debate about infants' 'belief' concepts might thus actually be terminological, and future discussion would much profit from more clarity as to which meta-theory of 'concept' is in use by different participants.

To summarize the discussion regarding R_2: For an observer to truly ascribe propositional attitudes to people, she has to fulfil some formal Generality Constraints: attitude ascription must be general in terms of contents and in terms of subjects of the attitudes. More data are needed in both respects to see whether infants fulfil such formal constraints. More specific constraints come into play when the attitude ascribed is supposed to be belief: in order to possess the concept 'belief', an ascriber must be able to track the functional and normative liaisons of beliefs to other mental states and to behaviour. Again, more data are needed to tell whether infants master such conceptual connections.

Summary and conclusion

The question 'Do infants have a theory of mind?' is currently much debated in ToM research. In fact, as it is treated by proponents and opponents of infant ToM capacities, it is not one clear question, but ambiguous between rather different readings, both regarding the representations of others' mental life infants use (first order) and the representations they represent others to have (second order).

Conceptually, the distinctions between levels of description and kinds of informational states reviewed here can help us detect, describe, and overcome such ambiguities and so avoid pseudo-debates between perfectly compatible claims phrased at different levels. Empirically, the conceptual distinctions help us focus the question and see what kind of empirical claims the data warrant. So, do infants then have a theory of mind? It depends. On a liberal reading, yes. Existing studies clearly show that infants have at least some subdoxastic states sensitive to some belief-involving situations. On a more stringent reading, however, couched in personal-level terms (Do infants have beliefs about beliefs?), we don't know yet. The existing data remain inconclusive in two respects at two levels: first, regarding whether what infants do in observing others is systematic enough (in terms of inferential integration, etc.) to count as involving beliefs proper; and second, regarding whether what they ascribe to others is general and specific enough to count as a belief.

Acknowledgements

Many thanks for helpful comments on an earlier version to Dana Barthel, Tanya Behne Steve Butterfill, Annette Clüver, Steffi Keupp, Josef Perner, Marco Schmidt, and Michael Tomasello. This work was supported by a 'Dilthey Fellowship' of the Volkswagen Foundation and the Fritz Thyssen Foundation and by the German Initiative of Excellence.

References

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118.

Bennett, M., & Hacker, P. (2003). *Philosophical foundations of neuroscience*. Oxford: Blackwell.

Bermudez, J. (2003). *Thinking without words*. Oxford: Oxford University Press.

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, *10*, 615–678.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*, 227–247.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337–342.

Carey, S. (2009). *The origin of concepts*. New York, NY, US: Oxford University Press.

Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge, MA: MIT Press.

Davies, M. (1989). Tacit knowledge and subdoxastic states. In A. George (Ed.), *Reflections on Chomsky* (pp. 131–152). Oxford: Blackwell.

Dennett, D. (1969). *Content and consciousness*. London: Routledge and Kegan Paul.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*(5), 735–808.

Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. New York, NY, US: Clarendon Press/Oxford University Press.

Gergely, G. (2011). The Development of Understanding Self and Agency. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 76–105). Oxford: Blackwell.

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowlege of intentionality. *Behavioral and Brain Sciences*, *16*(1), 1–14.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA, US: The MIT Press.

Koenig, M. A. (2002). Children's understanding of belief as a normative concept. *New Ideas in Psychology*, *20*(2–3), 107–130.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibly to others' beliefs in human infants and adults. *Science*, *330*, 1830–1834.

Kripke, S. (1972). *Naming and necessity*. Oxford: Blackwell.

Leslie, A. (2000). Theory of Mind as a mechanism of selective attention. In M. S. Gazzangia (Ed.), *The new cognitive neurociences* (pp. 1235–1247). Cambridge, MA: MIT Press.

Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, *9*(10), 459–462.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences*, *8*(12), 528–533.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J. (2010). Who took the "cog" out of cognitive science? Mentalism in an era of anti-cognitivism. In T. Frensch & R. Schwarzer (Eds.), *Perception, attention and action* (pp. 241–261). London, UK: Psychology Press.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*(2), 125–137.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, *308*(5719), 214–216.

Putnam, H. (1960). Minds and machines. In S. Hook (Ed.), *Dimensions of mind*. New York: New York University Press.

Putnam, H. (1975). The meaning of 'Meaning'. In K. Gunderson (Ed.), *Language, mind and knowledge* (pp. 131–193). Minneapolis: University of Minnesota Press.

Rakoczy, H. (2010). Executive function and the development of belief-desire psychology. *Developmental Science*, *13*(4), 648–661.

Rakoczy, H., & Tomasello, M. (2009). Done wrong or said wrong? Young children understand the normative directions of fit of different speech acts. *Cognition*, *13*(2), 205–212.

Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, *80*(4), 1172–1196.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.

Sellars, W. (1956). Empiricism and the philosophy of mind. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science, Vol. I: The foundations of science and the concepts of psychology and psychoanalysis* (pp. 253–329). Minneapolis: University of Minnesota Press.

Sellars, W. (1963). Some reflections on language games. In W. Sellars (Ed.), *Science, perception, and reality.* London: Routledge & Kegan Paul.

Shoemaker, S. (1968). Self-Reference and Self-Awareness. *Journal of Philosophy*, *65*, 555–567.

Shoemaker, S. (1981). Some Varieties of Functionalism. *Philosophical Topics*, *12*(1), 93–119.

Shoemaker, S. (1995). Moore's Paradox and Self-Knowledge. *Philosophical Studies*, *77*, 211–228.

Sirois, S., & Jackson, I. (2007). Social cognition in infancy: A critical review of research on higher-order abilities. *European Journal of Developmental Psychology*, *4*, 46–64.

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others referential communication. *Developmental Science*, *13*, 907–912.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief in two-year-olds. *Psychological Science*, *18*(7), 587–592.

Stich, S. (1978). Beliefs and subdoxastic states. *Philosophy of Science*, *45*, 499–518.

Strawson, P. F. (1959). *Individuals: An essay in descriptive metaphysics*. London: Methuen.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-olds. *Psychological Science*, *18*, 580–586.

Träuble, B., Marinovic, V., & Pauen, S. (2010). Early theory of mind competencies. Do infants understand false belief?. *Infancy*, *15*, 434–444.

Velleman, D. (2000). On the Aim of Belief. In D. Velleman (Ed.), *The possibility of practical reason*. Oxford: Oxford University Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749–750.

Received 23 December 2010; revised version received 12 August 2011