Jennifer Nagel – University of Toronto
January 19, 2017
Forthcoming in *Mind & Language*

Factive and non-factive mental state attribution

ABSTRACT: Factive mental states, such as knowing or being aware, can only link an agent to the truth; by contrast, non-factive states, such as believing or thinking, can link an agent to either truths or falsehoods. Researchers on mental state attribution often draw a sharp line between the capacity to attribute accurate states of mind, and the capacity to attribute inaccurate or 'reality-incongruent' states of mind, such as false belief. This article argues that the contrast that really matters for mental state attribution does not divide accurate from inaccurate states, but factive from non-factive ones.

1. Introduction

Research into mental state attribution focuses intensively on the capacity to attribute false beliefs. It is remarkable that we are able to grasp not only what others know about our shared environment, but also what others mistakenly take to be the case. Somehow, out of the countless ways in which an observed agent could be wrong about the world, we can identify just the right natural misconception and keep track of it, even as we also keep track of the divergent way in which reality itself is unfolding. On the basis of tests of explicit false belief attribution, it was once widely held that this capacity emerged around the age of four (Wellman, Cross, & Watson, 2001). On the basis of more recent research involving implicit measures, it has been argued that some capacity to recognize false belief appears much earlier, perhaps even as early as six or seven months (Baillargeon et al., 2014; Schneider, Slaughter, & Dux, 2015). The lag between implicit and explicit measures is a source of ongoing controversy, with some researchers arguing that there are two systems involved here (Apperly & Butterfill, 2009; Low, 2015), while others maintain that a single mindreading system faces obstacles of some kind in traditional explicit measures (Carruthers, 2016; Helming, Strickland, & Jacob, 2014; Leslie, Friedman, & German, 2004). Still others remain unconvinced that infants actually track false beliefs, arguing that key experimental results are explicable in terms of low-level perceptual novelty (Heyes, 2014), or by appeal to domain-general statistical learning processes, coupled with infant biases to attend to human faces, eyes, and behavior (Ruffman, 2014).

One benefit of this lively controversy has been increasing appreciation of the difficulty of accurate belief attribution. Even if mental state concepts are innate, we are not born knowing what others happen to

Acknowledgements: For comments and discussion, I am grateful to David Beaver, Peter Carruthers, Pierre Jacob, Hannes Rakoczy, Laurie Santos, Sergio Tenenbaum, Tim Williamson, Evan Westra, two anonymous referees for this journal, and to audiences at the 2016 *Society for Philosophy and Psychology* meetings, the 2016 *European Society for Philosophy and Psychology* meetings, MIT, Western University, William & Mary College, and the University of Indiana.
Address for correspondence: Department of Philosophy, 170 St George Street, Toronto, Canada
Email: jennifer.nagel@utoronto.ca

believe, and it is not easy to explain how we figure this out. Some systematic patterns apparently govern our intuitive calculations of what observed agents perceive, know, and believe to be the case, but it remains controversial just what these patterns are. Any theory of mental state attribution, whether the theory takes mental state concepts to be innate, early- or late-developing, faces a tough challenge in explaining how we are able to compute the contents of other agents' beliefs in a changing world, especially in cases where these agents are mistaken.

The depth of this computational challenge can easily be obscured by the tremendous ease with which we as adults attribute mental states (Heyes, 2014; Ruffman, 2014). We immediately and effortlessly see others as aware of certain features of the environment, as unaware of other features, as doing things intentionally, as trying, aiming and thinking, and it is natural for us to describe actions in these terms. In particular, when discussing the tasks taken to demonstrate false belief understanding in infants, we naturally describe the contrasting scenarios shown to the infants as involving agents with aims who see (or fail to see) that various things have happened. As Celia Heyes observes (2014, 648), it takes considerable artifice and effort to describe these scenarios without our familiar vocabulary of mental states and agency ('the arm-shape moved in an oblique, minimum jerk trajectory…'). However, when we are trying to figure out how (or whether) infants apply mental state concepts, it is important not to project our mature automatic registration of mental states onto the infant in the course of our description of the experimental setup. From the fact that an infant has seen an agent witnessing a change in her environment, it does not trivially follow that the infant has seen *that* this agent is witnessing this particular change. Theorists could of course maintain that certain mental state attributions are computed automatically from the start: for example, one could characterize infants as innately programmed to take observed agents to perceive (or think, or know) that salient events are happening, whenever agents are present. However, positing this type of innate programming imposes non-trivial burdens of explaining how such attribution principles might be genetically encoded and triggered by the infant's experiences.

Taking care not to smuggle in any unaccounted-for mental state computations, all theories of mental state attribution (or 'mindreading') ultimately need to solve the same problem. The developing mindreader observes persons, objects, and events, and computes, whether through innate or learned principles, or some combination of these, which mental states to attribute to whom, under various conditions of presence, absence, orientation and occlusion. From the perspective of an adult able to compute the contents of another's beliefs automatically, it is a challenge to examine the inner logic of these computations.

This article takes a fresh look at that inner logic. The aim here is not to take any position on whether genuine belief attribution starts early in infancy or later, but simply to examine the computations it involves. What does the competent mindreader need to see the observed agent as perceiving, knowing, and believing?

At a deeper level, what fixed relationships between conditions like perceiving, knowing, and believing enter into our calculations of these mental states? This article will discuss the relationship between perceiving and knowing, but its main focus is on the relationship between knowing and believing. Closer examination reveals a striking pattern in the contents of the beliefs that are tracked on standard false belief tasks: in what follows, I argue that these contents are always derived from the contents of prior attributed states of knowledge, and aim to explain why mental state attribution works this way. I begin with a closer look at the relationship between knowledge and belief.

The relationship between knowledge and belief is clearly a close one, and in the case of true belief, very close indeed. The standard epistemological view, going back to at least the 4th century BCE, is that knowledge entails true belief (Plato, c. 369 BCE/1990): if Smith knows that the barn is burning, then the barn must actually be burning, and Smith must believe that the barn is burning. The entailment is not mutual, however: there are cases of true belief that fall short of knowledge, for example in the ignorant person's accidentally true judgment about which road leads to Larisa (Plato, c. 380 BCE/1976). A wishful thinker could have a true belief that her lottery ticket will win, if she is lucky, without having known in advance of the draw. But if these examples illustrate that knowledge is a stronger condition than true belief, it nevertheless seems plausible that most cases of true belief whose formation we witness (notably ordinary perception in favorable circumstances) are also cases of knowledge. It might seem that there are no interesting differences between knowledge and true belief in the prediction of action, especially in the simple scenarios explored by developmental psychologists. If an observed agent is right that something is the case—if she has a true belief—why should it matter whether she furthermore meets the stronger condition of *knowing* that it is the case? When the agent wants something, then it might seem there is no difference between seeing this agent as having a true belief that it is in the basket, and seeing her as knowing that it is in the basket: either way, we expect the same immediate reaching behavior. True belief and knowledge are so closely related that it might seem the distinction between them could make no difference relevant to mental state attribution.

There is an approach to mental state attribution that minimizes the distinction between knowledge and true belief, employing a more generic conception of information across this divide. In the developmental literature, this route focuses on 'children's ability to attribute two different kinds of informational states to agents: *reality-congruent* and *reality-incongruent* states' (Song & Baillargeon, 2008, 1789). On this direct approach to the problem of false belief, false beliefs are a special type of informational state, sometimes also called 'counterfactual' (e.g. Baillargeon, Scott, & Bian, 2016), distinct in kind from informational states in which agents get it right. Both types of informational state are seen as spelling out the content of the attributed mental state, according to the following definitions: (1) 'Reality-congruent informational states specify what accurate information an agent possesses or lacks about a setting', and (2) 'Reality-incongruent informational states specify what inaccurate information an agent possesses about a setting' (Song & Baillargeon, 2008,

1789). Researchers in this program take their main challenge to be explaining the transition from attributing states of type (1) to the harder task of attributing states of type (2).

Prominent research in this tradition labels the contrasting scenarios that infants are responding to—for example, scenarios involving an observed agent who either does or does not witness an object's change of location—as *true belief* (TB) and *false belief* (FB) conditions (Onishi & Baillargeon, 2005). Actions are always seen as guided by beliefs, on this approach, combined with motivational states such as desires or goals, with true beliefs being easier to attribute than false beliefs, either because true belief is a more common and therefore default state (Leslie et al., 2004) or because an earlier-developing subsystem is responsible for true belief attribution (Baillargeon, Scott, & He, 2010).

It might seem that informational mental state attribution will be fully covered if we identify all states specifying the accurate and inaccurate information possessed by an agent; however, this approach leaves two gaps. First, it is not clear how it would classify attributions of beliefs whose truth value is unknown by the attributor ('he thinks that the box contains cereal, but I'm not sure if he's right'). Second, it is unclear how it handles attributions of knowledge which specify just the question whose answer is known to the agent, rather than the information constituting this answer ('he knows what is in the box—unlike me, he can see into it from where he is sitting'). Being able to see others as knowing the answers to questions enables us to represent the mental states of those who are epistemically better-positioned than we are: we do not have to be able to specify the accurate information they possess. These two gaps may raise doubts about whether mental state attribution is best modeled in terms of reality-congruent and reality-incongruent states, where the attributed true and false contents are specified.

As an alternative to drawing a line between true and false beliefs, or accurate versus inaccurate information, another approach would be to mark a fundamental split between factive and non-factive mental states, where factive mental states (like knowing) can link an agent only to truths, and non-factive states (like believing) link an agent to either truths or falsehoods. This approach has no problem with attributing beliefs of unknown truth-value: from the start, belief is conceived as a state of mind which may be held to either true or false propositions. True beliefs and false beliefs are not different types of mental state, but a single type of state coupled with different outward circumstances. Non-factive mental states like belief can retain their identity as external conditions change, so an agent's belief *that the ball is in the basket* is the same mental state, and is expected to produce the same action, before and after the unwitnessed transfer. Factive mental states like knowledge, by contrast, are more restricted: this kind of state by its very nature can link an agent only to the truth. These states are sharply restricted in how they may be formed and sustained, ceasing to exist if their contents change in truth value. When the agent fails to witness the ball being shifted out of the basket, her knowledge that it is in the basket is lost. Knowing is a strong condition to meet; section 3 argues that the

strength of this condition initially makes its presence and absence easier to track in other agents. One advantage of this approach is that it can explain why non-human primates respond differently to states of knowledge and states of true belief falling short of knowledge, a finding which is hard to explain if we lump these together as reality-congruent states. Finally, the truths that an agent knows can either be specified directly, using *that*-complements (the agent knows that the ball is in the box) or indirectly, using *wh*-complements (the agent knows where the ball is, the agent knows what is in the box); section 2 explores the significance of the factive/non-factive contrast in embedding questions. In summary, on this approach, whether we specify the content or the question whose answer is known, some actions are naturally seen as guided by knowledge, and others by mere beliefs (true or false), again in conjunction with motivational states.

These two ways of carving up the territory—the true/false belief way and the factive/non-factive mental state way—may seem like notational variants of each other. Indeed, advocates of the first approach are clear that knowledge is to be included as a reality-congruent state, and now sometimes label their contrasting conditions as knowledge versus false belief, rather than true belief versus false belief (e.g. Baillargeon et al., 2016, 174). Elsewhere they cash out true belief conditions as being conditions of knowledge, describing the unwitnessed transfer task as follows: 'True-belief scenarios served as control conditions and differed from false-belief scenarios in that the protagonist *knew* at the end of the trial sequence where the object was located' (Schneider et al., 2015, 3; emphasis added). The true/false belief way of approaching mental state attribution is distinguished not by any blanket refusal to speak of knowledge, but by a tendency to treat true belief and knowledge as roughly interchangeable, and by a demarcation of false belief as a special kind of mental state (e.g. 'counterfactual', as in Baillargeon et al., 2016). By contrast, on the approach advocated in what follows, false belief is not classified as a special type of mental state: it is rather just plain belief that counts as a mental state, a type of mental state that could be either true or false, depending on outward circumstances. Although knowledge entails true belief, it is not necessarily the case that a mindreader capable of recognizing knowledge is capable of recognizing true belief. This approach recognizes the difficulty of attributing false beliefs, but also highlights the difficulty of tracking of true beliefs that fall short of knowledge. It is not false belief, but belief as such that is harder to represent than knowledge.

Section 2 examines the contrast between factive and non-factive mental states in more detail. Although the terms 'factive' and 'non-factive' were coined in linguistics (in Kiparsky & Kiparsky, 1970), philosophers have adapted the linguistic classification of certain verbs as factive to forge a deeper classification of corresponding mental states as factive, with knowledge distinguished as the most general factive mental state (Williamson, 2000). Drawing on research from both disciplines, this section identifies key features of factivity relevant to the theory of mental state attribution.

Section 3 looks at the contrast between factive and non-factive mental state attribution in the tasks used with prelinguistic infants and nonhuman primates. Nonhuman primates appear to have some ability to recognize knowledge, but little if any capacity to recognize either false belief, or true belief that falls short of knowledge. Humans by contrast do at some point become capable of tracking the contents of true and false beliefs; this section looks at the initial calculation of those contents in our grasp of what others perceive and know. Section 4 examines minimalist efforts to explain the early stages of belief attribution with the help of notions such as 'registration' (Apperly & Butterfill, 2009), and 'experiential record' (Perner & Roessler, 2012). These can be seen as 'proto-factive' notions; at their point of origin, they can attach only to real objects and events, but they are not quite factive mental state notions, because their content is not propositionally structured. I argue that to the extent that these notions fall short of capturing factive mental states, they will not quite do the work needed to enable the emergence of belief attribution. Whether or not these notions can ultimately explain the level of success infants have shown on implicit false belief tasks, they leave a gap that needs to be filled in explaining how belief contents are calculated, whenever belief attribution begins. This is not just a problem for minimalists: the same gap arises even for those who take full-blown mental state concepts to be innate. I argue that factive mental state attributions play a pivotal role in initially determining the contents of the beliefs that are instinctively attributed to agents, and in guiding the ways in which those attributions are updated over time.

2. Factivity in linguistics and epistemology

One of the clearest lines of distinction between knowledge and belief concerns the relationship of each of these attitudes to the truth. The first of these sentences leaves the truth of its embedded proposition an open question, the second does not:

(1) Alice thinks that she is late.
(2) Bob knows that he is late.

We can go on to elaborate (1) by saying either 'Alice thinks that she is late, and she is right,' or 'Alice thinks that she is late, but she is wrong.' Sentence (2) is not so flexible: it sounds contradictory to say, 'Bob knows that he is late, but he is not late,' and it sounds redundant to insist on the truth of what is known. As a factive verb, *know* can embed only true complements; unlike belief, knowledge itself is an attitude that one can have only to truths. We can of course mistakenly attribute knowledge of a false proposition, but once the falsity of the proposition is recognized, the knowledge attribution must be retracted.

Table 1: factive and non-factive expressions

Factive	**Non-factive**
knows that	thinks that
is aware that	is sure that
realizes that	hopes that
is happy that	expects that
sees that	imagines that
recognizes that	is confident that
registers that	assumes that
notices that	believes that

Factive verbs such as *know*, *see* and *realize* entail the truth of their complement clauses. If Carla now sees that the ball is in the basket, it follows that the ball is in the basket; one cannot realize that one's wallet is missing unless it is actually the case that one's wallet is missing. Philosophers typically use the term 'factive' just to capture this property of entailment; as Timothy Williamson's now standard formulation puts it, 'a propositional attitude is factive if and only if, necessarily, one has it only to truths' (2000, 34). Among factive mental states, knowledge seems to have a special status: Williamson has influentially argued that it is distinguished as the most general member of this class, and indeed that it is entailed by all other members of the class. On this view, other factive mental states like *seeing that p* or *being aware that q* are simply more specific ways of knowing; these various factive mental states may differ from each other, so that for example *perceiving that p* does not necessarily entail *being happy that p*, but every factive mental state entails knowing (Williamson, 2000, ch.1).[1]

Linguists agree that *know* is factive, but use the term 'factive' more restrictively than philosophers; this article will follow the linguists in their more stringent usage.[2] Linguists reserve 'factive' for predicates that not

[1] Williamson's thesis that all factive mental states entail knowing is widely but not universally accepted; in particular, some philosophers have had doubts about emotive factives, like *regret* (see, e.g. Fantl 2015). The dispute is hard to adjudicate because the data on emotives are subtle, not least because these verbs easily invite a 'projected' or non-literal use (Holton, 1997). In addition, emotive factives have a number of unusual features, including a resistance to embedding questions (Egre, 2008): they take only declarative ('that-') complements, where most factives take interrogative ('wh-) complements (one can notice or know whether p, etc.). From a mindreading perspective, emotive factives will also involve complications about their combination of emotion with informational state. Emotive factives will therefore be set aside in what follows, and we will focus on regular cognitive factives such as *realize, register, see,* and *know*.

[2] A predicate of veracity such as 'is right that' will count as factive in the philosophers' sense, but not in the linguists' sense. Although such predicates link agents only to true propositions, they are not mental state expressions. One piece of evidence for this, noted by Pranav Anand and Valentine Hacquard (2014) is that these predicates do not select for sentient subjects: even a book or report can be right that something is the case, while we cannot say 'this report knows that…'. They argue that such predicates mark contributions to a common conversational ground, rather than mental

only entail but also presuppose the truth of their complements (Egre, 2008; Kiparsky & Kiparsky, 1970). What is presupposed is taken for granted by the sincere speaker as a background truth, in a way that does not shift even as the factive expression is embedded under a variety of logical and modal operators. Consider a simple knowledge attribution:

(3) Carla knows that the ball was moved.

Uncontroversially, (3) entails that

(4) The ball was moved.

Interestingly, background fact (4) also typically follows from all of the following sentences (Geurts & Beaver, 2011):

(5) Carla doesn't know that the ball was moved. *(negation)*

(6) Does Carla know that the ball was moved? *(question)*

(7) Perhaps Carla knows that the ball was moved. *(possibility modal)*

(8) If Carla knows that the ball was moved, she won't look for it in the basket. *(antecedent of conditional)*

If we substitute a non-factive verb like *thinks* for *knows* into (3), or any of (5)-(8), then background fact (4) no longer follows.

There are a number of rival theories of how exactly it is that presuppositions 'project' or shine through various types of embedding, but the differences between these theories will not concern us in what follows. All theories agree that it is possible to take special steps to cancel an embedded presupposition, or block the inference to the background fact—one could, for example, explicitly deny it, as in (9):

(9) Carla doesn't know that the ball was moved, because the ball wasn't actually moved.

The cancellation in the second clause brings with it a feeling that 'know' was the wrong word to have used here. When factive verbs appear unembedded, however, and in embedded contexts where no special steps are taken to cancel the natural inference, the use of a factive verb signals a commitment to the truth of its complement clause. In general, we use factive verbs both positively and negatively to mark and track the mental states of agents against a shared background reality: ordinary thought and talk of what people know and don't know simultaneously expresses our own grasp of that reality. If someone informs you that Diane knows that the flight arrives at Terminal C or that she is not aware it is raining, you can learn something not

states; remarkably, they observe, there seem to be no mental state predicates which entail their complements without presupposing them. Certainly, all the factive predicates in Table 1 are factive in the stronger sense.

only about Diane, but also about the airport and the weather.[3] Non-factive locutions, by contrast, offer no default delivery of information about their complement clauses.

Because the natural domain of what people do and do not know is just the background set of true propositions or facts, factives are essentially simpler to attribute than non-factives. When one makes the instinctive judgment that someone knows that *p* or fails to know that *p, p* itself is part of one's existing picture of the world. Factive projection patterns suggest that even questions about what people know, hypothetical reasoning about what they know, and speculation about what they might know, are naturally drawn against the background of reality. Meanwhile, the domain of what is merely believed is much less constrained, including true and false propositions alike. Various aspects of reality are known or unknown; the realm of what is believed is a vastly wider field. It is of course true that we can subsequently mark out the extension of the verb 'know' against that wider field, for any proposition and any agent: for example, having established which truths the agent S knows, one could then say of any proposition *p* outside this realm, true or false, that it is not the case that S knows that *p*. However, this is a somewhat deliberate and artificial exercise: the calculated external negation here contrasts with the ordinary sentential negation employed in sentences of the form 'S doesn't know that *q*', where *q* is simply presupposed to be true (Horn, 1985).

One last noteworthy feature that distinguishes verbs like *know* from verbs like *think* is their capacity to embed questions (Hintikka, 1975, ch 1). We may say not only that Carla knows that the ball was moved, but also that Carla knows whether the ball was moved, that she knows what was moved, and where it moved, and so forth; similar possibilities are open for many factives, including *notice*, *realize* and *see* (Lahiri, 2002). Cognitive non-factives like *think* and *believe* do not directly embed questions in this manner: it makes no sense to say that Carla thinks whether the ball was moved, or that she believes where it was moved. One key part of the explanation of this phenomenon is that the semantic contribution of the embedded question is the true answer to that question, so that verbs need to select for true complements in order to embed questions (Egre, 2008).[4] The person who knows where the ball is must be able to give the true answer to the question, 'Where is the ball?'. The person whose mental state about the ball is just thinking rather than knowing is not guaranteed to have the true answer here. Being able to embed questions (or *wh*-complements) under *know* and other factives enables us to report the mental states of those who are epistemically better positioned than we are: we can see others as experts on points that remain open questions for us. The question-embedding character of factives also enables us to represent weaker epistemic positions: we can say that Carla does not

[3] What the speaker presupposes about reality when using a factive verb will often be informative to others; indeed, studies of naturally occurring conversations suggest that more than half the time factives are used, information introduced as a presupposition of the factive will be new to the hearer (Spenader 2003).

[4] It is cross-linguistically robust fact that cognitive factives like *know* embed questions and cognitive non-factives like *think* do not, but it would be an oversimplification to conclude that all and only factives embed questions. For a detailed discussion of the exceptions, including emotive factives such as *regret* and interrogatives such as *wonder,* see Egre (2008).

know where the ball was moved. These representations do not need to specify any inaccurate information held by the agent.

While the examples reported in this section have been in English, the phenomena are cross-linguistically robust. The phenomena of presupposition and recursion (or embedding) are considered universal (von Fintel & Matthewson, 2008). The World Loanword Database, which covers 41 languages from all inhabited continents, lists words meaning *know* and *think* (both in the sense embedding a propositional complement) in every language (Haspelmath & Tadmor, 2009). There are some languages that have a dedicated verb for false belief, but this lexicalization is far from universal and appears not to confer any general advantage in mindreading (Shatz, Diesendruck, Martinez-Beck, & Akar, 2003; Tardif & Wellman, 2000). Verbs meaning *know* and *think*, meanwhile, not only seem to appear in all languages but are heavily used across languages, appearing among the thirty most common verbs in languages as diverse as Arabic (Buckwalter & Parkinson, 2014), Mandarin Chinese (Xiao, Rayson, & McEnery, 2009), and Russian (Sharoff, Umanskaya, & Wilson, 2014). The contrast in question-embedding behavior between verbs like *know* and verbs like *think* also cuts across languages (Egre, 2008; Lahiri, 2002). The simplest explanation for these regularities is that they reflect a common underlying conceptual structure, just as cross-linguistic regularities in grammatical evidential markings are taken to reflect a common underlying structure in human source monitoring capacities (cf. Papafragou, Li, Choi, & Han, 2007).

Factive verbs are striking in their prevalence, especially in children's speech.[5] A verb meaning 'know' figures heavily in children's usage: Bartsch and Wellman's (1995) study of English-speaking children up to the age of six found 'know' as the main verb in 70% of children's epistemic claims, with 'think' following at 26%; a similar study of Spanish-speaking children found 'know' (*saber*) leading mental verbs at 67% (even after formulaic 'I dunno' utterances were stripped out) and the two most common belief verbs (*creer, pensar*) together summing to 11% (Pascual, Aguado, Sotillo, & Masdeu, 2008). Paul Harris and colleagues have recently argued that this earlier work underestimates children's early competence in speaking of knowledge: against the common practice of discounting 'I dunno' statements, they have argued that 2-year-old's denials

[5] There is mixed evidence on children's early competence with these verbs in explicit mental state attribution. Negation may pose a difficulty: one study of three-year-olds found that they performed as well as adults in inferring *p* from sentences of the form 'S knows that p', but were significantly less successful in inferring *p* from sentences of the form 'S does not know that *p*' (Dudley, Orita et al. 2015). The significance of this finding remains unclear, not least because the pragmatics of the task are difficult, perhaps ambiguous, and require second-order mental state attribution: the child must infer that the toy is in the red box from hearing a hint in the form of a negated mental state attribution from an experimenter who should be taken to be aware of the location of toy. After whispering with puppet Lambchop, whose own statement is inaudible, the experimenter says, 'Lambchop does not know that the toy is in the red box.' Just 39% of the 3-year-olds gave the 'right' answer on this task, compared to 74% of the adults. The fact that the adults were so far from unanimity itself suggests that the task is open to several interpretations; it remains an interesting (and open) question how the children and adults understood it.

of knowledge (and their questions about knowledge) are not simply fixed conversational gambits but genuine mental state references, exhibiting clear sensitivity to what is known (Harris, Yang, & Cui, 2016). Harris and colleagues found strikingly similar patterns in talk of knowledge in English and Mandarin-speaking toddlers, a result consistent with earlier findings exploring the timing of factive and non-factive verb use. Factive verbs appear early, across languages: for example, use of a verb meaning 'know that' emerged before the verb meaning 'think that' in all of the Mandarin and Cantonese-speaking children in Tardif and Wellman's (2000) longitudinal study. They take their findings to bolster Anna Wierzbicka's (1992) proposal that words for *want, know,* and *think* are universal features of human language—these terms figure in her very short list of 'lexical universals'—suggesting that 'children worldwide acquire these terms and meanings in a clear order,' with talk of wanting followed by talk of knowing, and only later by talk of thinking (Tardif & Wellman, 2000, 35).

3. Mental state attribution in humans and other primates

Belief attribution is difficult, and one piece of evidence for this difficulty comes in comparing the performance of humans to great apes, and to more distantly related primates. Humans clearly form robust representations of the true and false beliefs of others. Great apes may perhaps show some glimmerings of false belief understanding: in unwitnessed transfer tasks, they show some tendency to look in the correct direction as they anticipate the action of a deceived agent (Krupenye, Kano, Hirata, Call, & Tomasello, 2016), although their performance on these tasks is perhaps explicable in terms of domain-general 'submentalizing' mechanisms (Heyes, 2016). Great apes do not seem to represent the false beliefs of other agents robustly enough to guide action: in competitive food choice tasks, they fail to exploit a competitor's natural misconception about hidden food (Hare, Call, & Tomasello, 2001; Kaminski, Call, & Tomasello, 2008). Moving to more distantly related primates, monkeys have shown no signs of any capacity to represent false belief, even in looking-time experimental paradigms closely modeled on those taken to show implicit false belief recognition in infants (Marticorena, Ruiz, Mukerji, Goddu, & Santos, 2011; Martin & Santos, 2014).

Knowledge attribution appears to be easier: great apes act in ways that are consistent with differentiating knowledge from ignorance, for example in paradigms where they need to compete with a dominant animal who should be seen as either knowledgeable or ignorant of the location of hidden food, given that animal's observed track record of perceptual access (Hare et al., 2001). Even monkeys seem to be able to distinguish between knowledgeable and ignorant competitors (e.g. Flombaum & Santos, 2005).[6]

[6] Alia Martin and Laurie Santos (2016) have raised questions about whether monkeys are able to represent ignorance. They observe that 'the act of representing a relation between an agent and a piece of information that is not part of our current reality is a computational challenge,' and suggest that monkeys lack 'the capacity to conceive of states of the

There is evidence that nonhuman primates may have some ability to attribute not only knowledge-*that*, but also knowledge-*wh* (Krachun, Carpenter, Call, & Tomasello, 2009). In a task involving a search for hidden food, chimpanzees watched as a human competitor observed food being placed in one of two out-of-reach opaque containers. The chimpanzee's view was obstructed so that the chimpanzee did not have first-hand knowledge of the location of the food, but the chimpanzee could see that the human had a clear line of sight to the containers during baiting. In the false belief conditions of this experiment, the two containers were switched while the competitor left the room momentarily or turned away. At the moment of test, when the chimpanzee was allowed to approach the containers, the competitor then reached unsuccessfully for one of them, so the subject's reaching for the other container would be consistent with a grasp of false belief: chimpanzees did not pass this test. In the 'true belief' (or more intuitively, knowledge-where) condition of the experiment, however, the containers were swapped after baiting in plain view of both the chimpanzee and the knowledgeable human competitor, so the chimpanzee's reaching for the same container would be consistent with a grasp of the competitor's knowledge. Chimpanzees performed significantly better than chance on this version of the task: although they had not seen the placement of the food themselves, and so could not 'specify the accurate information possessed by the agent' in their initial representation of the competitor's mental state, they followed the behavior of the knowledgeable agent in their choice of container, as if guided by a representation of that agent as knowing where the food was.

As further evidence for the greater ease of knowledge- over belief-attribution, there is evidence than nonhuman primates have difficulty representing true belief that falls short of knowledge (Kaminski et al., 2008, Study 2). In another food choice task, chimpanzees competed with conspecifics for food hidden in opaque containers. Subject chimpanzees could choose a high-quality food reward they and their competitor had seen being hidden in one of three containers; these containers were placed on a table which could slide back and forth between the subject and the competitor, a table which was always offered to the competitor first. Subjects also had the option of a low-quality fallback reward in a fourth container. Subjects had to make a single choice from the containers after their competitor had chosen (and perhaps emptied) a container, but they could not see their competitor's choice, so they needed to make an inference about that choice based on their calculation of the competitor's mental state. In the simplest conditions, after the initial baiting, both the subject and the competitor watched as the high-quality reward was either lifted from one container and replaced there ('known lift') or lifted from one container and moved to another ('known shift'). In the more challenging conditions, after the initial baiting, the subject chimpanzee could see that the competitor's view of the table was briefly occluded by a screen as the reward was either lifted and replaced ('unknown lift') or lifted and transferred to another container ('unknown shift'). Chimpanzees were significantly more likely to choose

world that are different or 'decoupled' from their own current reality.' (p.376) However, instinctive attributions of ignorance do not involve such decoupled representations: they involve real-world facts that the ignorant agent is missing.

the container with the high-quality reward when the competitor had not witnessed the reward's final motions, consistent with taking the knowledge or ignorance of their competitor into account. However, they failed to distinguish between the unknown shift (false belief) and unknown lift (true belief) conditions: they treated their competitor as simply ignorant in both. Six-year-old children did better: they were most likely to choose the high-quality reward container in the unknown shift condition, expecting the competitor to have the false belief that the reward was still where the competitor had seen it last. Unlike chimpanzees, children were significantly less likely to pursue the high-quality reward in the 'unknown lift' condition, where it was replaced in the same spot where the competitor had initially seen it, anticipating that the competitor would retain a true belief about the reward's location.

Evidence of difficulty in grasping true belief falling short of knowledge has also been reported for Rhesus monkeys (Santos, 2016). These monkeys also ordinarily anticipate that a returning agent will reach for a target object where they saw it being hidden; however, when this object is publically removed from that location in the agent's absence, they no longer show any expectation about where the returning agent will reach for it, even if the object is restored to its original hiding place prior to the agent's return. It is as if the attributed state of knowledge or awareness of location, a state that would be maintained for a stationary object during an agent's absence, or updated if the object is moved as the agent watches, is shattered when the object is moved behind the agent's back. When the object moves unseen, the agent's attitude towards the location of that object is no longer of a kind that can only be held towards truths; the agent ceases to know where the object is, and cannot regain that state of knowledge by having the object returned to its original location as her view is obstructed. Knowledge is a state with rigorous entry conditions.

One source of difficulty in building a theory of mental state attribution is in identifying its triggering conditions. Hannes Rakoczy characterizes the problem as follows: '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' (2012, 71). Indeed, in the case of *false* beliefs, we have something harder than an absence of superficial features: the attributor who is witnessing a scene in which *p* is the case must override some aspects of what presently meets the eye in order to attribute to another the inaccurate belief that not-*p*. Yet the possibility of inaccuracy is widely agreed to be an essential part of propositional attitude ascription, Rakoczy reports: 'From the beginning of ToM research, there has been a wide-ranging consensus that what are needed to demonstrate the use of propositional attitude concepts are tasks that require a subject to understand that agents represent the world from their specific subjective points of view—that is, potentially differently from the ways in which others represent it, and potentially inaccurately' (2015, 2). The prelinguistic mindreader faces a steep challenge, if this wide-ranging consensus is correct.

However, taking factive attitudes as our starting point makes the challenge more tractable. Because factive propositional attitudes like knowledge can only get things right, a potential for inaccuracy is not essential to propositional attitude concepts. Factive attitude ascriptions also show that the two aspects of subjectivity identified by Rakoczy are separable: agents can be seen as knowing more or less than the attributor, and therefore as representing the world differently, without necessarily representing it inaccurately. I can see you as failing to know what is in the container without seeing you as having an inaccurate representation of its contents; I can also see you as knowing more than I do on some issue, thanks to your superior vantage point. Attributions of knowledge and ignorance capture subjectivity in the sense of ascribing different perspectives on reality to different subjects, without going all the way to a stronger sense of subjectivity, in which the subjective perspective is actually, or even potentially, at odds with what is objectively the case.

Could there be specific superficial features initially associated with the class of knowledge-involving situations? Indications of perceptual access are natural candidates here, with typically developing infants becoming increasingly sensitive over time to observable cues such as the direction in which others are facing, their direction of gaze, and the presence of occlusions to lines of sight (for a review, see Moore, 2008). If visual access, for example, is a matter of *seeing that something is the case*, then it is a matter of having a factive mental state (alternative accounts of perceptual access will be discussed in due course). Factive mental states are promising entry points of mental state attribution because in these basic nonverbal tasks the attributor needs to make sense of an interaction between the agent and the environment, and some salient fact in that environment needs to figure in setting the content of the attributed mental state. The triggering conditions for attributing perceptual access are not taken to leave it an open question whether the attributed content is true or false; the kind of state being attributed—awareness of some event or feature of the environment—is a kind which could only have true contents. Conditions marking the absence of knowledge are also easily detectable: when someone's view is obstructed, or her back is turned, she can easily be recognized as ignorant, where it will be a more difficult question what beliefs she may possess.

One might wonder how to reconcile a factive account of perceptual access with the possibility of misperceptions: given an appropriately placed colored filter, an observed agent could form a mistaken belief about an object's color, for example, and the attributor who sees the filter between the agent and the object could naturally understand this. From the attributor's perspective, however, the attribution of a mistaken representation of the object's color will be a more complex process than attribution of normal perceptual access. The attributor who can see the object and the agent's line of sight to that object through the colored filter can readily attribute to him perceptual access to the object's presence, but she will need to infer or calculate his misrepresentation of the object's color on the basis of her own past experience of the filter as

well as what meets the eye in the moment. The agent can see a blue object *as* green, when there is an interfering yellow filter, but one cannot *see that* an object is green if it is not.

So-called 'false perceptions' do not count as moments of perceptual access, on the factive account of perceptual access. This point is worth emphasizing, because the terminology is potentially confusing, tempting us to think of 'false perceptions' as members of a broader kind ('perceptions') which should be taken to exclude them, if 'perceptions' are understood in the ordinary way, as moments of perceptual access to reality. Care with these labels does not necessarily challenge the main conclusions researchers wish to draw, but enables us to avoid some of their more problematic claims, and indeed to highlight key features of their arguments (bracketing, for now, rival lower-level explanations of the phenomena). For example, increased terminological caution is useful in interpreting the results of Hyun-Joo Song and Renée Baillargeon's (2008) exploration of apparent false belief tracking in 14.5-month-old infants. In a violation-of-expectation study, the infants watched an agent repeatedly demonstrate a preference for a doll with long blue hair over a toy skunk. These items were then placed in opaque boxes either in the presence or the absence of the agent. The box in which the skunk was hidden had a potentially deceptive tuft of blue hair protruding from under the lid (the 'hair box'); meanwhile, the blue-haired doll was placed in an unmarked box (the 'plain box'). In the condition where the agent was present during the placement of the toys (the 'true perception' condition, arguably better classified as a knowledge condition), after the blue-haired doll was placed in the plain box, infants reliably looked longer if the agent reached for the hair box, as if surprised that the agent was not reaching for the desired doll at its known location. By contrast, in the condition where the agent was absent during the placement of the toys, infants looked longer if the returning agent reached for the plain box, although this was the accurate location of the desired doll. This response was taken to support the conclusion that 'the infants had to reason that the agent's *false perception* of the tuft of hair as a part of the doll would lead her to hold *false beliefs* about the locations of the doll and skunk' (2008, 1794). Here the attribution of a 'false perception' of the tuft of hair *as* a part of the doll is itself already a false belief attribution: if these researchers are right, the infant must attribute to the returning agent the false belief that this tuft of hair is part of the doll, from which, as they say, a false belief about the doll's location would follow and be expected to guide action.

In the reasoning that the researchers attribute to the infants, attributions of the non-factive state of *perceiving as* are not introduced on the same level as attributions of factive states of knowledge or perceptual access; indeed, these attributions of the non-factive state of *perceiving as* are performed only subsequently to, and based on the contents of, prior attributions of factive states. The initial stages of the infants' mental state reasoning are properly characterized as operations on factive states: 'The infants also had to consider the agent's knowledge of the setting in the test trial: They had to attribute to the agent not only the ability to *detect* the tall boxes and tuft of hair, but also the ability to *infer* that the doll and skunk were likely to both be

present, as in the preceding trials, and to be hidden in the boxes' (2008, 1794). Here, the initial acts of detection constitute perceptual access in the strict sense: detection is a relation that can only hold to real objects. If the infant left it as an open question whether or not the agent was registering the presence of the tuft of blue hair on the outside of the hair box, there would be no basis for the infant's expectation in the false belief condition that the agent would reach for that box. As a factive attitude, seeing or detecting *that* there is a tuft of blue hair on the hair box is a process which is different in kind from perhaps mistakenly seeing this tuft *as* a part of a doll: a capacity to attribute the first, factive kind of state must be in place to enable one to attribute the second, non-factive kind of state.

The false beliefs that are supposed to be attributed in standard false belief tasks are closely related to prior knowledge attributions. The unexpected transfer task is the simplest: at some earlier time, the agent is taken to know the location of an object, having seen it there. Ordinarily, turning away briefly from a stationary inanimate object does not destroy one's knowledge of its location: knowledge is a state which is triggered by perceptual access, but a state which typically endures past that initial access. Having seen the ball in the basket at a given time, one continues to know that it is there a moment later, even without continuing observation. Indeed, it is because knowledge can outlast perceptual access that knowledge attribution has value in predicting behavior: if states of knowledge did not endure through time, there would be little advantage in positing a deeper state behind the agent's observable marks of perceptual access. The moment the agent turned away, the content of what she had witnessed would have no impact on her future behavior: simple behavior reading and detection of sightlines could take the place of mental state attribution for others aiming to anticipate her actions. It is because knowledge of location is typically retained over time that an intelligent agent will reach for the object where she saw it last, and a successful mindreader will anticipate this knowledgeable behavior. When the ball moves behind the agent's back, a primate incapable of representing belief simply registers a failure of knowledge and comes to see the agent as ignorant, no more likely to search one place than another. To represent belief, one needs not only to see the agent as ignorant of the object's new location (her back was visibly turned while it was moved), but also as retaining her tendency to act as she did when she had knowledge of where the object was. Believing is on this view a shadow or after-effect of knowing: the deceived agent who reaches for the basket behaves as if she knew that the ball is in the basket, while her state of mind is no longer latched onto the truth. Knowledge that the ball was in the basket initially degenerates into mere belief that it is still there at a slightly later time, when an agent is ignorant of the shift.

In an unexpected contents task, the false proposition that the agent should be seen as believing ('there are candies in this box') does not start out as known by the agent. Rather, this false belief is attributed to the agent only following attributions of appropriately related knowledge and relevant ignorance. First, the agent is seen as seeing or knowing that there is a particular type of container before her: this successful apprehension of some aspect of reality is vital input to the mindreader's calculations. The observed agent is

also seen as ignorant of its contents, not having seen inside when the contents were revealed to the attributor, and therefore as inferring, on the basis of her background knowledge about this kind of packaging, that it contains its typical contents; again, belief is an after-effect of what is known. The blue-haired doll task is structurally similar: the returning agent is taken to see or know that there is a tuft of hair protruding from the box, and to infer the doll's location from this fact, together with her background knowledge about the intact doll's appearance.

Success on misinformation tasks is also enabled by factive mental state attribution. In a standard misinformation task (Perner & Wimmer, 1988), the agent is taken to know what has been said, and to form a false belief on this testimonial basis. The successful mental state attributor cannot leave it as an open question whether the agent has heard the misleading testimony, or whether the agent has taken a trusting or skeptical attitude towards what she has heard. Insofar as the mindreader sees the agent as inferring that *p* is the case from being told that *p,* the mindreader is interpreting the agent as following our general schema for accepting others' testimony as knowledgeable, arguably a schema built into our ordinary norms of conversation (Grice, 1975). Because the agent is misinformed in this case, the mindreader is of course not ultimately attributing to her the knowledge that *p,* but he attributes to her the belief that *p* on the basis of the attributed knowledge that *p* has been said to be the case, together with an expectation that the agent will treat this testimony as knowledgeable.

Although the field of propositions that can be believed is larger than the field of propositions that can be known, it is not intractably larger, at least in these basic tasks. Unexpected transfer tasks leave agents believing things that very recently used to be the case, and were known; unexpected contents tasks leave agents believing things that would ordinarily be the case, based on what these agents are seen as knowing about salient objects or object parts in their present environment, and misinformation tasks leave agents believing things that they know have just been said to be the case, by informants the agents take to be knowledgeable. A key element in all of these tasks is the agent's observed perceptual access. If the target agent were not taken to have perceptual access to the object's original location, to the deceptive packaging, object part, or testimony, there would be no reason to attribute the particular false belief that is naturally attributed to the agent. The mindreader who attributes a false belief does so on the basis of witnessing an observed agent's interactions with reality, and needs to start her calculations about the character of this belief with input from what that agent has taken up, and carry them forward with input about the agent's ignorance, or the aspect of reality which are blocked for her: in these basic tasks, factive mental state attributions launch and guide attributions of non-factive states.

If Williamson is right that all factive mental states entail knowing, then seeing that something is the case entails knowing that it is the case: the agent who sees that the ball is in the basket must know that the

ball is in the basket. However, even if we assume that there is factive mental state attribution here, it remains a live question whether the developing mindreader who is cued by observable signs of visual access would attribute to the agent the more generic mental state of *knowing that p* or the more specific factive mental state of *seeing*—in the sense of 'knowing through current visual access'—*that p*. There are some reasons to suspect that it is initially the more generic state of knowing that is attributed: notably, the epistemic state generated by seeing is taken to outlast visual access to the scene. Even when the agent no longer sees that the ball is in the box, because the lid has been closed or it is well inside it, out of a direct line of sight (as in Onishi and Baillargeon 2005), the agent is still taken to know that the ball is in the box; if in early development it is typically the more generic state of knowing that is encoded from the start, the infant can take this state to guide the agent's later performance without any further generalizing or inferential steps. Knowledge is a relatively enduring underlying state, on this picture, whose attribution is triggered by signs of perceptual access without being hostage to the continued presence of those signs. Attribution of the more generic state of knowledge would also make it easier to integrate the perceptual and inferential or testimonial components of unexpected contents and misinformation tasks. As somewhat more direct evidence for early attribution of knowing-that as opposed to just seeing-that, even at age three, children are somewhat weak at tracking the particular sensory modality responsible for what an agent knows (O'Neill, Astington, & Flavell, 1992; Papafragou et al., 2007). Meanwhile, successful early task performance is not limited to tasks in which the agent has gained knowledge through sight: for example, 15-month-olds react in a way consistent with knowledge attribution to an agent who manipulates the position of a unseen ball by tilting a ramp behind her back (Träuble, Marinović, & Pauen, 2010). Further research on perceptions of knowledge acquisition through a greater variety of modalities would help to clarify whether more specific or more generic factive mental states are first attributed, while also helping to resolve the age at which any type of genuine mental state attribution begins.

4. Perceptual access, minimalism, and the challenge of computing beliefs

Even if it is uncontroversial to take perceptual access as the starting point for epistemic or informational mental state attribution, it is controversial whether or when attributions of perceptual access should be understood as attributions of knowledge as opposed to belief, or more radically, some non-propositional relation. On the two-systems model, infants are capable of a minimal form of mindreading, which tracks relationships between agents, objects and locations. Representations of these relationships—'belief-like states'—function as intermediaries between the infant's input of environmental cues concerning an agent and the anticipated behavior that the infant expects on the part of this agent, without representing perceptions or propositional attitudes as such (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). To each observed

agent, on this theory, the infant assigns a *field* of objects, where the scope of the field is defined by factors such as orientation, lighting and occlusions; an agent is seen as *encountering* an object if it is within her field, and as *registering* this object as located at the place last encountered. Infants expect goal-directed action to be governed by registration, where 'registration is like belief in that it has a correctness condition which may not obtain: a registration is correct when the object is in the location [where the agent most recently encountered it]' (2013, 617). It is a delicate question how attributions of these belief-like states fall short of being belief attributions. While one might expect that the correctness condition on registration already qualifies it as a propositional attitude—accepting the mainstream view of propositions as above all, 'the primary bearers of truth and falsity' (King, Soames, & Speaks, 2014, 5)—Ian Apperly and Stephen Butterfill insist that registration is non-propositional in character. In their view, the key mark of the propositional is its capacity to represent a state of affairs under a particular perspectival aspect: to attribute propositional attitudes at all, the mindreader must be able to distinguish between various ways in which a single object might be identified (for example, as the Morning Star or as the Evening Star). If registration just links agents to objects, rather than to possibly mistaken representations of objects, it should not, in their view, count as propositional. They argue that infant mindreading has the signature limits predicted by this view: for example, an incapacity to track false beliefs about the identity of an object which has multiple aspects, such as a robot which is blue on one side and red on the other, potentially mistaken for two different robots by agents who have moments of suitably restricted perceptual access to it (Low & Watts, 2013). Signature limits are also posited for other minimalist theories of infant social sensitivity, such as Josef Perner and Johannes Roessler's view, in which the infant simply keeps an 'experiential record' of the objects and events that have recently engaged the observed agent. When an agent who was distracted or absent returns to a scene, this record is reanimated, producing responses of surprise at any novelty, relative to this record, in the agent's behavior, just as a function of unfamiliarity rather than as a product of any genuine mental state attribution or reasoning about what the agent should rationally do (Perner & Roessler, 2012). The experiential record is supposed to be short-lived, non-propositional in character, and unlikely to integrate with the child's desires to produce deliberate action.

Claims about signature limits can be tested, and both of these minimalist programs have felt some heat. As one of its authors acknowledges (Perner, 2014), in response to a line of criticism pressed by Peter Carruthers (2013), the experiential record view is hard-pressed to explain the results of 'helping paradigm' experiments, in which an infant helps an agent unlock one of two boxes, in a way which seems to reflect whether the infant takes the agent to have a true or false belief about the contents of the box (D. Buttelmann, Carpenter, & Tomasello, 2009). These experiments call for deliberate action rather than signs of surprise, where the infant's action seems to be guided by some relatively enduring representation of the agent's epistemic position. Carruthers also argues that the tasks intended to show a failure to represent object identity

cannot be taken as direct support for minimalist views of mindreading, because these tasks call not just for mindreading, but also for executive abilities beyond the infant's range, such as the ability to visualize a 180-degree rotation of an object like the two-color robot (2013). Meanwhile, in simpler tasks involving objects with ambiguous aspects (such as a sponge which looks like a rock), 18-month-old infants seem to be capable of helping adults selectively in a way consistent with their tracking the agent's grasp of one or another aspect of the object (F. Buttelmann, Suhrke, & Buttelmann, 2015).

If experimental results challenging minimalism prove robust, the discovery of behavior consistent with propositional attitude attribution may still not end the debate about whether this behavior is in fact driven by propositional attitude attribution. For many past experiments, minimalists have crafted ingenious low-level explanations of the infants' response patterns (Heyes, 2014; Ruffman, 2014). There is continuing debate about the viability of those particular explanations, given various control condition results, and deeper debate about which style of explanation is more parsimonious, about the extent to which parsimony even matters, and about the relative burdens of proof on minimalist and mentalist explanations (Ruffman, 2014; Scott, 2014).

The line between minimalist and mentalist accounts can be hard to draw. Perner makes a useful observation on this frontier in his (2014) response to Ted Ruffman's defense of minimalism. He observes that minimalism faces an ongoing challenge in differentiating itself from mentalism, as minimalists develop more sophisticated stories about how much infants can learn from observing agent perceptual access and behavior patterns. When Ruffman endeavors to explain infants' expectations of how agents will act, given those agents' observed track records of perceptual access, Perner notices that Ruffman slips into talking about infants' expectations of what the agent *should* do, given the facts she has witnessed, and Perner is concerned that here 'the model defined by the facts of what the agent has witnessed comes dangerously close to being treated as a belief' (2014, 296). Perner does not exempt himself from this worry: looking at his own theory of experiential records (Perner & Roessler, 2012), and Apperly and Butterfill's (2009) theory of registration, he finds a fundamental commonality: 'Both notions are based on distinguishing events that a person has experienced (registered, encountered, witnessed, seen, …) from those events that the person hasn't witnessed. (…) The facts *registered* by the agent are supposed to be taken by the infant as the basis for calculating the agent's behavior, which in effect makes the registered facts to function as the content of the agent's belief about the current state of the world' (2014, 296). It is noteworthy that if even minimalist programs end up with something akin to belief contents ('belief-like states'), they derive these contents from some type of factive condition, a condition such as encountering or witnessing, of a type that can only link an agent to a fact or real state of affairs. When the facts originally registered no longer obtain, the agent can still be governed by some conception of those earlier facts (whether these are propositional in character or just encode clusters of agents and objects). Figuring out how this later model or conception will emerge from the

earlier facts is a non-trivial task; indeed, whether we ultimately need a richer or leaner understanding of perceptual access (or witnessing, or encountering) to explain performance at any given stage of development will be determined in part by what is needed to develop an account of how contents are updated through the transition between original access and later deployment.[7] Propositional structure enables sophisticated forms of updating and integration with background knowledge in inference, and in thinking about questions. Lean accounts will be challenged for a given stage of development if for example, we gain robust evidence that mindreaders at that stage reliably succeed at non-verbal knowledge-where tasks of the container-switch type, where the knowledgeable agent is not seen together with the target objects at the moment of baiting.

Such tasks also challenge accounts of perceptual access which see propositional structure as attributed from the start, while taking these attributions to be non-factive in character. Carruthers' single-system innatist position is in this category. He takes the following line against minimalism:

> There seems no reason to think that the early mindreading system is incapable of attributing propositional thoughts to other agents. On the contrary, since infants *have* propositional thoughts from the outset themselves (as Apperly, 2011, acknowledges), they can take whatever proposition they have used to conceptualize the situation seen by the target agent and embed that proposition into the scope of a 'thinks that' operator. (Carruthers, 2013, 162)

'Thinks that' is a controversial choice for the operator here. If infants get mental state attribution off the ground by embedding their own propositional grasp of a real-world situation seen by the agent under a mental state operator, then it is more natural to characterize this operator as a factive one, like 'knows that'. From the perspective of the theorist, it is not a mistake to describe the knowing agent as also thinking that something is the case—knowing does after all entail believing—but the infant would be throwing away information by encoding the observed agent as having only belief. Seeing infants as starting with belief attribution rather than knowledge attribution also creates a puzzle about how children then regain enough information about which belief states constitute knowledge to support their early facility with talk of knowledge. Meanwhile, knowledge attributions already have special practical value in prelinguistic tasks: if I take an observed competitor to see and therefore know which of two containers holds a reward, I have reason to follow her subsequent choice, a reason I would lack if I merely see her as having some belief on this matter.[8]

---

[7] For a detailed argument that the contents of registrations end up needing to be managed in a way that calls for propositional structure, see Jacob (forthcoming).

[8] Belief-centered accounts are challenged to explain cases where the information possessed by the agent is not directly available to the onlooker. For example, Agnes Kovács, who takes the fundamental representational structure in mindreading to be a 'belief file', suggests that the attributor opens an 'empty belief file', which she describes as a file in which the content remains 'undefined'. However, in explaining the inferences mindreaders can make in these cases, she herself characterizes them as having a capacity "to recognize that some other person *knows* where the object is while we ourselves do not know where it is" (2016, 518). It does seem more promising to describe these structures as embedding questions, rather than as empty files.

The focus of this paper has been on mental state attribution tasks used with pre-linguistic infants and nonhuman primates. Creatures with language no longer have to watch how an agent acts upon objects to attribute beliefs to her: with language, they can attribute beliefs on the basis of what the agent says, including what she says about absent or nonexistent objects, or what clashes with the hearer's grasp of reality: representations 'decoupled' from reality are now immediately available as input to their calculations. The capacity to attribute mental states to other agents on the basis of their assertions does not necessarily wipe out the importance of factive mental state attribution, if, for example, our default in understanding those assertions is to take the speaker as knowing, or if the transmission of knowledge plays a key role in the structure of assertion (Williamson, 2000, ch 11). Assertions would have little value if we generally took them to be radically decoupled from reality. Still, given the common potential for gaps between reality and what agents say, and given the general power of belief-desire explanations in predicting human action, we might expect to see adults generally switching to the weaker concept of belief—a state of mind which may or may not match reality—to account for what agents do, perhaps with some default presupposition that beliefs tend to be true. However, we do not seem to give up on the concept of states of mind whose essence involves capturing the truth: factive mental state attributions continue to play a prominent role in adult mental state attribution. Whether because it is less taxing to map out an agent's state of mind against reality, as opposed to a larger space of propositions, or because there are some types of action that are better explained by appeal to knowledge than by appeal to belief (Nagel, 2013; Williamson, 2000, ch 1), or for some other reason, we retain the idea of a type of mental state which could only attach to the truth, and references to this type of state continue to play a large role in adult explanations of behavior. In mundane social navigation, we typically see each other as acting for reasons that are anchored in facts in the world, rather than as driven by internal states that are potentially decoupled from reality (Perner & Roessler, 2012). But even when agents' internal states of belief end up being at odds with reality, instinctive tracking of those states is guided by our larger picture of intelligent behavior and agency where this is first understood in terms of the agent's interaction with our shared environment: predictably forming certain false beliefs is one of the many things we see agents do as they interact with the real world. The line between factive and non-factive mental states is therefore an important line for theorists of social intelligence to watch.

References

Anand, P., & Hacquard, V. 2014. Factivity, belief and discourse. In L. Crnic & U. Sauerland (Eds.), *The Art and Craft of Semantics: A Festschrift for Irene Heim* (Vol. 1, pp. 69-90). MIT: CreateSpace Independent Publishing Platform.

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

Baillargeon, R., Scott, R. M., & Bian, L. 2016: Psychological reasoning in infancy. *Annual Review of Psychology, 67*, 159-186.

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

Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K.-s., . . . Bian, L. 2014: Psychological and sociomoral reasoning in infancy. *APA Handbook of Personality and Social Psychology: Vol1 Attitudes and Social Cognition. Washington, DC: APA*.

Buckwalter, T., & Parkinson, D. 2014: *A frequency dictionary of Arabic: Core vocabulary for learners*: Routledge.

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.

Buttelmann, F., Suhrke, J., & Buttelmann, D. 2015: What you get is what you believe: Eighteen-month-olds demonstrate belief understanding in an unexpected-identity task. *Journal of Experimental Child Psychology, 131*, 94-103.

Butterfill, S. A., & Apperly, I. A. 2013: How to construct a minimal theory of mind. *Mind & Language, 28*(5), 606-637.

Carruthers, P. 2013: Mindreading in infancy. *Mind & Language, 28*(2), 141-172.

Carruthers, P. 2016: Two systems for mindreading? *Review of Philosophy and Psychology, 7*(1), 141-162.

Dudley, R., Orita, N., Hacquard, V., & Lidz, J. 2015. Three-year-olds' understanding of know and think *Experimental Perspectives on Presuppositions* (pp. 241-262): Springer.

Egre, P. 2008: Question-embedding and factivity. *Grazer Philosophische Studien, 77*(1), 85-125.

Fantl, J. 2015: What Is It to Be Happy That P? *Ergo, an Open Access Journal of Philosophy, 2*.

Flombaum, J. I., & Santos, L. R. 2005: Rhesus monkeys attribute perceptions to others. *Current biology, 15*(5), 447-452.

Grice, H. P. 1975: Logic and conversation. *Syntax and Semantics, 3*, 41-58.

Hare, B., Call, J., & Tomasello, M. 2001: Do chimpanzees know what conspecifics know? *Animal Behaviour, 61*(1), 139-151.

Harris, P. L., Yang, B., & Cui, Y. 2016: 'I don't know': Children's early talk about knowledge. *Mind & Language*.

Haspelmath, M., & Tadmor, U. (2009). World Loanword Database. Retrieved from http://wold.clld.org/

Helming, K. A., Strickland, B., & Jacob, P. 2014: Making sense of early false-belief understanding. *Trends in cognitive sciences, 18*(4), 167-170.

Heyes, C. 2014: False belief in infancy: a fresh look. *Developmental Science, 17*(5), 647-659.

Heyes, C. 2016: Apes Submentalise. *Trends in cognitive sciences, 1629*.

Hintikka, J. 1975: *The Intentions of Intentionality*. Dordrecht: D. Reidel.

Holton, R. 1997: Some telling examples: A reply to Tsohatzidis. *Journal of pragmatics, 28*(5), 625-628.

Horn, L. R. 1985: Metalinguistic negation and pragmatic ambiguity. *Language*, 121-174.

Jacob, P. forthcoming. Challenging the Two-Systems Model of Mindreading. In A. Avramides & M. Parrott (Eds.), *Knowing and Understanding Other Minds*. Oxford: Oxford University Press.

Kaminski, J., Call, J., & Tomasello, M. 2008: Chimpanzees know what others know, but not what they believe. *Cognition, 109*(2), 224-234.

King, J. C., Soames, S., & Speaks, J. 2014: *New thinking about propositions*: Oxford University Press.

Kiparsky, P., & Kiparsky, C. 1970. Fact. In M. Bierwisch & K. Erich (Eds.), *Progress in Linguistics* (pp. 143-173). The Hague and Paris: Mouton.

Kovács, Á. M. 2015: Belief Files in Theory of Mind Reasoning. *Review of Philosophy and Psychology*, 1-19.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. 2009: A competitive nonverbal false belief task for children and apes. *Developmental Science, 12*(4), 521-535.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. 2016: Great apes anticipate that other individuals will act according to false beliefs. *Science, 354*(6308), 110-114.

Lahiri, U. 2002: *Questions and answers in embedded contexts*. Oxford: Oxford University Press.

Leslie, A., Friedman, O., & German, T. 2004: Core mechanisms in theory of mind. *Trends in cognitive sciences, 8*(12), 528-533.

Low, J. 2015. Two-Systems View of Children's Theory-of-Mind Understanding. In R. Scott & S. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences*: Wiley.

Low, J., & Watts, J. 2013: Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science, 24*(3), 305-311.

Marticorena, D. C. W., Ruiz, A. M., Mukerji, C., Goddu, A., & Santos, L. R. 2011: Monkeys represent others' knowledge but not their beliefs. *Developmental Science, 14*(6), 1406-1416.

Martin, A., & Santos, L. R. 2014: The origins of belief representation: Monkeys fail to automatically represent others' beliefs. *Cognition, 130*(3), 300-308.

Martin, A., & Santos, L. R. 2016: What cognitive representations support primate theory of mind? *Trends in cognitive sciences, 20*(5), 375-382.

Moore, C. 2008: The development of gaze following. *Child Development Perspectives, 2*(2), 66-70.

Nagel, J. 2013: Knowledge as a mental state. *Oxford Studies in Epistemology, 4*, 275-310.

O'Neill, D., Astington, J., & Flavell, J. 1992: Young children's understanding of the role that sensory experiences play in knowledge acquisition. *Child Development, 63*(2), 474-490.

Onishi, K. H., & Baillargeon, R. 2005: Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255.
Papafragou, A., Li, P., Choi, Y., & Han, C. 2007: Evidentiality in language and cognition. *Cognition, 103*(2), 253-299.
Perner, J. 2014: Commentary on Ted Ruffman's "Belief or not belief:…". *Developmental Review, 34*(3), 294-299.
Perner, J., & Roessler, J. 2012: From infants' to children's appreciation of belief. *Trends in cognitive sciences, 16*(10), 519-525.
Perner, J., & Wimmer, H. 1988: Misinformation and unexpected change: testing the development of epistemic-state attribution. *Psychological Research, 50*(3), 191-197.
Plato. c. 369 BCE/1990: *The Theaetetus of Plato* (M. J. Levett, Trans.). Indianapolis: Hackett.
Plato. c. 380 BCE/1976: *Meno* (G. M. A. Grube, Trans.). Indianapolis: Hackett.
Rakoczy, H. 2012: Do infants have a theory of mind? *British Journal of Developmental Psychology, 30*(1), 59-74.
Rakoczy, H. 2015: In defense of a developmental dogma: children acquire propositional attitude folk psychology around age 4. *Synthese*, 1-19.
Ruffman, T. 2014: To belief or not belief: Children's theory of mind. *Developmental Review, 34*(3), 265-293.
Santos, L. R. 2016. *The evolution of theory of mind: Insights from non-human primates*. Paper presented at the Society for Philosophy and Psychology , 42nd Annual Meeting, Austin, Texas.
Schneider, D., Slaughter, V. P., & Dux, P. E. 2015: What do we know about implicit false-belief tracking? *Psychonomic Bulletin & Review, 22*(1), 1-12.
Scott, R. M. 2014: Post hoc versus predictive accounts of children's theory of mind: A reply to Ruffman. *Developmental Review, 34*(3), 300-304.
Shatz, M., Diesendruck, G., Martinez-Beck, I., & Akar, D. 2003: The influence of language and socioeconomic status on children's understanding of false belief. *Developmental Psychology; Developmental Psychology, 39*(4), 717.
Song, H.-j., & Baillargeon, R. 2008: Infants' reasoning about others' false perceptions. *Developmental Psychology, 44*(6), 1789.
Spenader, J. 2003: Factive presuppositions, accommodation and information structure. *Journal of Logic, Language and Information, 12*(3), 351-368.
Tardif, T., & Wellman, H. M. 2000: Acquisition of mental state language in Mandarin-and Cantonese-speaking children. *Developmental Psychology, 36*(1), 25.
Träuble, B., Marinović, V., & Pauen, S. 2010: Early theory of mind competencies: do infants understand others' beliefs? *Infancy, 15*(4), 434-444.
von Fintel, K., & Matthewson, L. 2008: Universals in semantics. *Linguistic review, 25*(1/2), 139.
Wellman, H., Cross, D., & Watson, J. 2001: Meta analysis of theory of mind development: The truth about false belief. *Child Development, 72*(3), 655-684.
Williamson, T. 2000: *Knowledge and its Limits*. New York: Oxford University Press.