

RADU J. BOGDAN

WATCH YOUR METASTEP

The First-Order Limits of Early Attributions

—

Abstract: There is a wide and puzzeful gap between the child's mastery of first- and recursive or higher-order attributions of attitudes, measured not only in years but also in the cognitive resources involved. Some accounts explain the gap in terms of the maturation of the competencies involved, others invoke the slow development of enabling resources, such as short-term memory, the syntax of sentence embedding or sequential reasoning. All these accounts assume a continuity of competence between first- and higher-order attributions. I disagree and argue, with psychological and neuroscientific support, that there are two distinct (though developmentally overlapping and interacting) competencies, one metaintentional and the other metarepresentational. I focus below on the former and argue that it is egocentric, situated, nonpropositional and thus intrinsically limited to first-order attributions, even when all the enabling resources are in place.

1. INTRODUCTION

There are several intriguing puzzles about the phylogenetic and ontogenetic emergence and operation of second- and higher-order attributions of attitudes, such as beliefs about beliefs. Here is a short list. On the one hand, we are told that apes and in particular chimpanzees can do higher-order attributions. Also, according to the standard Gricean account of communication, which nowadays dominates the field, very young children should be able to communicate by making higher-order attributions of attitudes, such as recognizing intentions to induce beliefs. On the other hand, there is increasing skepticism about any sort of ape interpretation, not just higher-order, and, on the human side, we are told that children develop second-order attributions rather late, perhaps no earlier than the age of six, and it takes them several more years to elevate this ability to still higher orders, even though they can interpret simple intentional relations, such as seeing or wanting, since infancy and more complex relations, such as belief, as early as three to four.

Again, on the one hand, we are told that higher-order attributions require no more than first-order attributions plus enabling resources of a general or multipurpose computational sort, such as a capacious short-term memory, the syntax of sentence embedding and sequential reasoning. On the other hand, such enabling resources are largely available by six or seven and yet children still cannot manage higher-order attributions, particularly those involved in telling stories about people, concocting or detecting lies, producing and understanding humor, passing blame or responsibility, making moral judgments, or evaluating commitments. Add to these puzzles the disturbing discrepancy in estimates about the timing of higher-order attributions -- from the teens (Flavell et al. 1968) to the six to ten interval (Perner 1988) or even five years and a half (Leekam 1991) -- and we get a volatile if not incoherent picture of recursive or higher-order interpretation.

I do not have a solution to these puzzles nor do I propose a novel account

of attitude attributions in general. But I do think that these puzzles, together with some developmental clues, invite a rethinking of first-order versus higher-order attributions. In this paper I limit myself to showing that first-order interpretation cannot do higher-order attributions, not even when all the standard enabling resources are developed and utilized, for reasons that distinguish it, as a competence apart, from higher-order interpretation.

A few words about how I propose to proceed. In the next section I argue that the earlier and simpler forms of interpretation, documented in apes and young children, have features, such as egocentricity, situatedness and a bias for concrete or objectual (as opposed to propositional) contents, that enable them to represent only simple intentional relations to concrete items and prevent them from representing complex attitudes to propositions and, as a consequence, from handling higher-order attributions. Against prevailing views, section 3 argues that the recognition of false belief is in the same conceptual boat and, to that extent, still short of being propositional, metarepresentational, and capable of higher-order exploits. Sampling neuropsychological data, section 4 confirms the distinction between first-order metaintentional interpretation and higher-order metarepresentational interpretation by showing that the former is done mostly on the left side of the brain, whereas the latter relies massively on the right-side of the brain. Section 5 draws some implications.

It may be already apparent that I treat the word 'interpretation' as synonymous to 'theory of mind' or 'naive psychology.' I construe the interpretive relation to obtain between an interpreter and not just a subject (the individual interpreted), let alone his mind, but rather the subject's relation to the world. To a very significant extent, differences in phylogenetic and ontogenetic forms of interpretation emerge from differences in which aspects of subject-world relations are represented by the interpreter and how. It is such differences that bear on whether a form of interpreting other minds can or cannot go beyond first-order attributions.

2. METAINTENTIONALITY

Apes as well as young human children interpret others in ongoing or present-tense contexts of perception, motivation and action, exclusively from a subjective or egocentric perspective. I group these features under the notion of situated interpretation (a notion inspired by Perner's 1991 situation theory of mind). Situated interpreters represent observable subject-world relations, such as seeing or noticing, or relations easily read from behavioral symptoms, such as wanting or attending.

The situated recognition of a subject's intentionality consists in categorizing and tracking three elements -- the relatedness of the subject to something in the world, the direction of that relatedness and its target as a concrete and spatio-temporally defined item, be it object, event or situation. The recognition of relatedness (more exactly, a purposed or active relatedness) may originate in a naive-biological representation of animacy and agency or goal-directedness, a representation later appropriated by naive psychology or interpretation and woven into its conceptual gadgetry (Bogdan 2000; 2001). The representation of the direction of agency and hence relatedness appears to rest mainly on the ability to detect bodily orientation and gaze. Serviced by a brain-specialized mechanism, this is a specialized interpretive ability that tracks subject-world relations and is apparently possessed only by higher primates (Baron-Cohen 1995). Although great apes are thought to represent relatedness and its direction, there is growing skepticism about their ability to identify the target involved (Povinelli 1996; Tomasello and Call 1997). They appear to identify the target of gaze or bodily orientation egocentrically, in terms of their own perception and motivation and not through a specialized and context-invariant skill of interpretation.

This tripartite representation of intentionality makes conceptual sense

because the three categories -- relatedness, direction and target -- do constitute our basic sense of intentionality or aboutness. It also makes empirical sense because interpreters may possess some of the categories but not others. As just suggested, apes may represent relatedness and its direction but not, qua interpreters, a target. Human infants begin by representing relatedness but represent its direction later and its target still later. Indeed, as argued later, the incompleteness of their schemes of interpretation, as well as their egocentric situatedness, is a key factor that prevents such schemes from handling higher-order attributions.

I propose to call metaintentional a tripartite scheme of interpretation (relatedness + direction + target) that operates in current contexts of perception and motivation, hence situatedly, and individuates targets (whether egocentrically or through a specialized skill) as concrete and spatio-temporally defined items (things, events, situations). On this construal, metaintentional interpretation represents a subject's intentional relations to concrete items (objectual attitudes, as logicians put it), as opposed to representing his attitudes to propositions. The latter scheme is metarepresentational, usually unsituated and mind-sensitive, and develops only in humans, probably after the age of five or six.

To illustrate the first-order limits of metaintentional interpretation, I will help myself to a synoptic reconstruction of some anecdotes found in the literature on ape interpretation (sources cited at the end of this section). Recall that chimpanzees and very young human children are thought to recognize the relational character of gaze and its direction but not its target. They appear to identify the target egocentrically, according to their own perception and motivation. Imagine now a situated metaintentional interpreter A (ape or very young child) who sees B watching C's gazing at some target X. A recognizes that B gazes at C, who gazes at an X that is visually shared by all the parties concerned. To count as metaintentional, gaze interpretation requires the

recognition of relatedness, its direction and target. The question is how these recognitions are achieved. If A recognizes B's gazing at C's gazing at X, isn't A representing B's directed relatedness to [target =] C's directed relatedness to [target =] X? If so, isn't A doing second-order metaintentional interpretation? In other words, isn't A representing intentional facts (or perhaps propositions) about B's representation of intentional facts (propositions) about C's representation of target X? I would say no, for several reasons.

To begin with, there is a scarcity of documented real-life situations in which nonhuman primates or very young children engage in sequentially connected gazings or other ongoing intentional relations, despite an abundance of social interactions involving three or more individuals. The abilities to track multiple intentional relations may simply not be there, particularly given the egocentric and item-sensitive or objectual (as opposed to propositional) grasp of targets. More generally, from the fact that A recognizes B's gaze and its direction toward C, it does not follow that A recognizes that the target of B's gaze is C's gaze and more specifically C's gazing at X. A may simply notice B gazing at C but not at C's gazing at something specific. A may notice B's gaze and its direction but determine its direct target, C, egocentrically, in terms of A's own perception of C. The same may be true of C's gazing at X: it may be determined by A (and also B) egocentrically, through perception. Noticing a subject's gaze could act as an instruction to look for a target through one's own perception, relative to one's goals and interests in a context.

Two further observations back this estimate. The first is that since situated agents live in the here and now of perception and motivation, they cannot inhibit what they see and want in order to leave room for and figure out what others see and want. (Inhibition plays a crucial role in the recognition of false belief and later in the development of unsituated and nonegocentric interpretation.) Yet, for vital reasons, they must figure out what others see and want. How do they do it? Most likely, by a division of labor between specialized

skills that track relatedness (such as agency) and its direction (manifested in gaze or bodily posture), on the one hand, and their egocentric and self-motivated perception of targets, on the other hand. Looking at later developments in human childhood, this division of labor makes sense because the determination of the target of a subject's intentional relations is a very difficult task, involving major innovations, such as joint attention, that elude both apes and human infants (Bogdan 2000, ch. 5; 2001; Tomasello 1999, chs. 3 and 4).

The other supporting observation is that in the world of nonhuman primates and very young children such intentional relations as gaze, wanting, noticing or attending are interpreted as directed toward concrete and spatio-temporally defined types of items, such as things, events and situations. Gazing at an individual, who also happens to gaze at something, is not likely to change the type identity of targets, from concrete items to intentional facts or propositions about gazers. The imputation of higher-order gazing assumes that an interpreter A represents an intentional fact about a subject B's relation to the world -- namely, B seeing that [target] and what fills the target slot is a second intentional fact about another individual's (C) gaze, namely, that C gazes at X. Yet a situated metaintentional interpreter of gaze (or wanting or attending) is limited to representing only one or two parameters of intentionality (relatedness and its direction). The missing parameter, the target of a subject's intentionality, is actually the one that would allow the embedding of another intentional relation. Yet the value of that parameter is determined egocentrically, by the interpreter's perception, and not by a specialized skill, and takes the form of a concrete item, not of a proposition, which alone would be conceptually open to include other propositions.

It is important to insist that even when a specialized interpretive skill tracks the target of a subject's intentional relation, the resulting (first-order) representation need not be that of proposition and therefore need not take the

form ‘subject sees or desires that ---’. The interpretive representation, usually perceptual, could still be objectual or item-oriented, hence metaintentional, as it is in joint attention and situated (false) belief. The advent of proposition in the interpretation of the target of an intentional relation appears to be a late ontogenetic achievement, perhaps as late as five to six, ushering in a radically new category, that of propositional attitude, and a new form, unsituated and metarepresentational, of interpretation.

If so much is plausible, then we cannot talk of a higher-order metaintentional interpretation of gazing or seeing (or wanting or attending). Yet my sense is that some authors are ready to countenance such a higher-order reading of metaintentional interpretation (e.g, Cheney and Seyfarth 1990, ch. 5; Whiten and Byrne 1988) and that there is wide sympathy for such a reading in the psychological as well as philosophical literature (see Whiten and Perner 1991, for a survey). The assumption behind this higher-order optimism may be that, after all, metaintentional interpreters do determine targets nonegocentrically, by specialized interpretive skills. Perhaps but not likely. And, as just noted, this possibility need not turn a metaintentional interpretation into a higher-order enterprise anyway. The upshot is that young children’s interpretation of gaze, desire and belief can be fully metaintentional and still fail to sponsor higher-order attributions because of its situated take on intentional relations and its focus on concrete, spatio-temporally defined items as targets of those relations. If defensible, this diagnosis would go some way toward explaining why children master first-order attributions of intentional relations years before they master higher-order attributions of propositional attitudes.

3. FALSE BELIEF -- A METAINTENTIONAL CATEGORY

The standard false-belief test has a child interpreter watching or being told the following scenario: a subject places an object X at a location A, then

leaves the room for a few minutes, during which, unbeknownst to him, the object is moved and hidden at a new location B; the subject returns to the room and the interpreter is asked to predict where the subject will look for X. Children younger than three choose the new location B and only children older than three indicate the old location A.

It is widely thought that by passing the false-belief test between the ages of three and four, children graduate to a radically new form of interpretation, which is supposed to be unsituated, because it transcends the confines of current perception and motivation; metarepresentational, because it captures a content that depends on how the subject represents it; and propositional, because (among other things) truth or falsity are values of propositions. This view, held by many if not most psychologists of interpretation (e.g., Astington et al. 1988; Lewis and Mitchell 1994; Perner 1991; Wellman 1990; Whiten 1991), entails that passing the false-belief test is conclusive evidence that the four-year-old has mastered the interpretive category of propositional attitude. There are developmental psychologists who think that this category may emerge even earlier in childhood, as evidenced by the mastery of pretense (Leslie 1987).

I propose to challenge this dominant view and its implication about propositional attitudes by arguing that the recognition of false belief is still metaintentional in all crucial respects, hence situated, limited to intentional relations and their concrete targets, and therefore not yet an unsituated metarepresentation of belief as propositional attitude. There are several reasons why passing the false-belief test need not transcend the powers and limitations of metaintentional interpretation. For starters, there is the dissident view that children younger than three might recognize false belief (Chandler 1988; Lewis 1994), in which case, on my analysis, they must do it in metaintentional and situated terms. There is also corroborating evidence that children are implicitly aware of false belief by looking at the correct location where the subject

believes an object is, before they can answer questions about that location (Clements and Perner 1994). This also suggests an implicit metaintentional and nonlinguistic solution to the false-belief problem.

Finally and most importantly, recognizing false belief is recognizing cognitive mismatch. Yet cognitive mismatch is not the only sort of mismatch that very young interpreters recognize situatedly and metaintentionally. Andrew Meltzoff (1995) has shown that 18-months-olds recognize conative mismatch as well, in the form of failed intentions or (more realistically said) failed attempts to reach some visible goal. Children younger than three also appear to recognize misaimed desires (Wellman 1990, chs. 8 and 10). With the help of alternative representations of actual and imagined situations, of which young children are capable (Perner 1991, ch. 3), failed tryings, misaimed desires and false beliefs could all be represented by the very young interpreter as a subject's relations to concrete items (objects, situations or events) -- in the conative case of tryings and desires (to put it graphically), as an arrow going from inside out but failing to reach their targets, and in the cognitive case of false belief, as an arrow going from outside in but failing to reach the subject's mind and cause his behavior, because of a past arrow that earlier did reach his mind and should cause his behavior. To sum up, the fact that children younger than three recognize mismatch generally in situated and metaintentional terms suggests that they can do it without resort to metarepresentation, that is, without attending to the mental specifics of how subjects' intentional relations fail to reach their targets, and without interpreting those intentional relations as relations to propositions.

It may already be apparent that I see a deep complicity between metarepresenting attitudes and having an interpretive grasp of propositions as targets of attitudes. To clarify this point, three caveats are in order. First, from the fact that a child understands and uses a propositional language, with fully formed sentences, it does not follow that the child has an interpretive grasp of

propositions as targets of attitudes, anymore than the child's linguistic usage of 'I think' need indicate an understanding of thought as mental representation or of thinking as operating over such representations (Flavell et al. 1995; Perner 1988). As noted in the next paragraph, the use of an interpretive language is not the same competence as interpreting attitudes.

Second, through language and her own noninterpretive cognition, the child interpreter may project propositions as targets of a subject's intentional relations but that would still be an egocentric projection, whereby those propositions are represented according to the child's mind, not according to the subject's, hence not interpretively. Differently said, it is the young interpreter's general cognition and language use, not her specialized competence for interpretation, that impute such propositions as targets of the subject's intentional relations. This imputation does not yet amount to a metarepresentation of propositional attitudes. In strict interpretive terms, the child interpreter still views the subject related to concrete items, hence metaintentionally. The child interpreter may think and say that a subject believes that the object is in the old place but actually interpret just the subject's memory of the object at the old location. In the same vein, even if the three-to-four-year old applies the grammar of sentence-embedding to attitudinal verbs such as 'desires' or believes' and hence understands or uses a phrase like 'Sam believes that Lucy believes that ...', it need not follow that her interpretive semantics of such phrases either places propositions in the scope of an attitudinal verb or else treats those propositions the way the adult does. Up to the age of three to four, the interpretive semantics of the child is very likely to be imagistic and concrete, hence metaintentional (in my sense).

This reading of the situation brings me to a third caveat, which concerns the variety of notions parading under the label of proposition. Truth and falsity can be global, unstructured values of propositions, as they are in propositional calculus. This unstructured notion of proposition is not the one needed to

explain the interpretive category of an attitude to a proposition. The notion needed is that of a proposition understood intensionally -- that is, understood according to how the subject is interpreted to represent it mentally, logically and often even grammatically. This is when metarepresentation, *stricto sensu*, comes into the ontogenetic picture. I find it significant that it is only around the age of seven (the age of composite propositional attitudes, such as intention or hope, and of second-order attributions) that children are able to interpret opaque or intensional attributions, which are sensitive in a fine terminological, logical and conceptual grain to how a subject represents something (Russell 1996, 219).

If, despite these clarifications, one still insists that the four-year child somehow interprets false belief and other forms of mismatch propositionally, then one should better envisage only the first, unstructured sense of proposition (Bogdan 2000, 136-137; see also Perner 1991, 120, 280-282). But, so far as I can see, that sense collapses into the category of a concrete item (thing, event, situation) as target of an intentional relation. This is to say that the metaintentional interpretation of (false) belief simply registers that the subject's knowledge of an object at a location is 'off target' (rather than 'on target') or is 'on old target' (rather than 'on new target') and nothing structurally more complex than that. For the four-year-old interpreter, the subject's mind counts minimally (just epistemically), as having or not having information about a target, without a sense of how the information may be represented. This is why the interpretation of false belief is not metarepresentational.

But, one may still wonder, could a child interpret belief at all without metarepresenting attitudes to propositions? I think she could. First of all, the child's notion of belief may be modeled on earlier interpretive categories of simple desire and perception or informational access (Wellman 1991) and the latter are situated and metaintentional. Mental development is likely to be

conservative and resort to old tricks whenever possible, before venturing in new directions. It is worth noting that children younger than three could not pass a test involving conflicting desires, which preserved the false-belief narrative but replaced belief with desire, construed metaintentionally or objectually, as in 'John wants water' (Russell 1996, 229-231). I read this as a failure of inhibition.

What the three-to-four-year-old interpreter of false belief seems able to do is inhibit (a) her current perception of object X at location B (that is, her own knowledge of the current situation) and also inhibit (b) a propensity to relate the subject to X, and (c) interpret the subject's past perception of X at the old location A (that is, the subject's only relation to X and hence his actual knowledge of X's location) as the real cause that explains the subject's behavior. Yet the subject's past perception is likely to be memorized and interpreted by the three-to-four year old in metaintentional terms. As far as I can tell, there need be nothing metarepresentational in this achievement, revolutionary as it may be in other respects. In other words, to ascertain the subject's false belief, the child interpreter need not (and probably does not) attend to how the subject actually represents object X at location A. Entertaining a subject's past intentional relation and what is now the case is not yet metarepresenting anything, since both are perception-based, even though such entertaining may signal an increased sensitivity to the subject's mental take on things.

I would propose the same diagnosis for the child's understanding of her own false belief. In a seminal experiment, Gopnik and Astington (1988) asked children what is in a box of Smarties. They answered 'Smarties'. The box was opened and it contained just pencils. Asked later what they thought the box contained when first shown to them, the answer of three-to-four year-olds was pencils. Children of this age and younger have as much difficulty with remembering their own past false beliefs as they have with predicting someone else's false belief. The difficulty, as I see it, is due first of all to their inability to

inhibit and override their current knowledge (of pencils) in order to allow access to a memory of what they first took to be in the box. The problem is not memory of past thoughts, as independent tests confirm (Perner 1991, 188). And the problem is not the ability of youngsters to entertain alternative representations of actual and possible (hypothetical, counterfactual, future) situations; they do this from a very early age. As noted earlier, this is probably how they handle mismatch. The problem, I think and suggest next, is contrasting a subject's currently perceived intentional relation with an alternative representation of another such relation (e.g., memory of a subject's past perception), overruling the former and interpreting in terms of the latter. Only inhibition allows such a contrast and overruling. These distinctions are worth spelling out.

Representing past situations (in memory) or entertaining alternative situations (in imagination) need not involve representing intentional relations. Before the age of three to four and before inhibition, children can do the former but not the latter. Why? My guess is that inhibition alone allows (a) a full disengagement from what is currently perceived and apt to be (mistakenly) interpreted and (b) the application of interpretive procedures to past or alternative situations populated by subjects (and selves) in intentional relations to sundry targets of interest. A full disengagement matters because the memory and (counterfactual) imagination of young children most often seem to be cued by what they currently perceive and want, including the currently interpreted intentional relations. This, in turn, could be a constraint that prevents them from resorting to alternative interpretations in past or counterfactual situations. Once freed from this constraint by inhibition, around the age of three to four, children can turn their existing interpretive procedures to past or counterfactual situations. In the case of false belief, around the age of three to four, those interpretive procedures are likely to be situated procedures that check perceptual access in past or counterfactual situations.

This is why, novel as it is in other respects, the resulting interpretation remains metaintentional, hence not metarepresentational, not propositional and not unsituated.

To sum up, then, what I think the false-belief attribution requires an interpreter to do is (at least) to

- disengage but hold in mind alternative and often conflicting metaintentional representations
- override a current representation of a target and inhibit the propensity to relate the subject intentionally to that target
- recover a memory of a subject's past intentional relation and treat that relation as causally involved in the subject's behavior
- interpret and predict in terms of the latter intentional relation

Among several competing accounts of false belief in the psychological literature, there is a lively debate between executive-function and interpretational accounts. Some discussions explicitly compare the merits of these major accounts (Perner and Lang 2000; Russell 1996; Zelazo et al. 1999). These accounts need not be incompatible. Alan Leslie combines inhibition (or selection processing, as he calls it) with interpretation (Leslie 2000; Scholl and Leslie 2001), a position also adopted here. There are, however, at least two differences between Leslie's account and mine. One difference is that, like many other researchers, Leslie analyzes false belief as a metarepresentation of a propositional attitude, whereas I analyze it in simpler metaintentional terms. The other difference is that, unlike Leslie, I am not convinced that prior to the inhibition-produced category of false belief, the young child actually understands belief in terms that are different from her earlier understanding of mere information (usually perceptual) access or lack thereof. If the interpretive (ToMM) module is designed always to attribute true

beliefs (Scholl and Leslie 2001, 697), that cannot provide an understanding of false belief; in which case what contributes to the understanding of false belief, namely inhibition, cannot simply be an improvement in performance. In my view, inhibition helps installing a new interpretive category. It is a complex but still metaintentional category and is made (at least) of the following elements:

false belief =

- (current) belief

[perhaps modeled on the prior interpretive categories of perception or desire as intentional relations to concrete items]

- that is not true = off target

[the interpreter's knowledge of the current situation + inhibition of this knowledge as interpretive avenue + memory of subject's past intentional relation, made interpretively salient]

As an executive development, inhibition contributes to the assembly or at least availability of a novel knowledge structure by juxtaposing two distinct and conflicting representations of intentional relations. Both are needed to produce the recognition of false belief. We may assume that after the age of four, with the new false-belief category firmly in place, recognizing an instance of false belief amounts most of the time to bringing it under this structure, with the help of relevant evidence. The resulting category is unlikely to be innate qua unitary and ready-made knowledge structure, as may well be the case with earlier metaintentional categories, such as those of gaze or desire or perhaps joint attention. Even if inhibition is triggered by innate mechanisms around the age of three to four, it is not a development within the child's knowledge of the world or within her interpretive competence, although it mightily impacts both. If later interpretive categories also owe a good deal to inhibition and other

nonconceptual developments, as I think they may, then their assembly cannot have an innate interpretation-dedicated blueprint, either.

This sketchy account of the category of false belief is essentially hybrid: it views the category as assembled out of several developments, executive as well as interpretive. It is consistent with a number of studies that have shown that children younger than three to four have difficulties handling competing representations of evidence, whether in handling false belief or distinguishing between appearance and reality or remembering their past beliefs or switching perspectives (Bjorklund 2000, 216-217). In most if not all of these cases, inhibiting the cognitively salient information on some current intentional relation is the major obstacle. Although inhibition is only one major player in the assembly of the category of belief and of later categories (such as opinion, intention or hope), its role is quite revolutionary in allowing the child interpreter to escape the situatedness of current perception and entertain and network diverse and often conflicting representations of intentional relations, without which she could not represent complex mental states. Grounded in robust neurological facts, particularly the late development of the inhibitory frontal lobe, appeal to inhibition has the advantage of providing a simple and economical explanation of a variety of significant changes in the child's mind around the age of three to four across a variety of domains. The representation of false belief is just one of these changes, in the domain of interpretation. It is an accomplishment which, through inhibition, opens the way to a new interpretive competence but which, by still operating metaintentionally, remains anchored in the earlier competence.

Some puzzles noted in the introduction bring further support to the notion that, difficult and innovative as it surely is, the child's first grasp of false belief need not transcend the limits of metaintentional interpretation. One puzzle was this. If around four, children interpret belief and other intentional relations as (allegedly) full-fledged propositional attitudes and also master the

syntax of sentence embedding, which they do, why do they need several more years to do higher-order attributions and engage in the complex interpretive practices, such as gossip, humor, or moral evaluation, that depend on such attributions?

It may be, as often thought, that children also need to mature other enabling computational resources, such as short-term memory and sequential reasoning, in order to do higher-order attributions successfully. The assumption here is that once children understand false belief and other intentional relations, and can embed sentences in other sentences, no further conceptual development within interpretation is needed to do higher-order attributions -- a position explicitly stated in Sullivan, Zaitchik and Tager-Flusberg (1994) and popular amongst many other psychologists and philosophers (Fodor 1992; Leslie 1987). For reasons that will soon become apparent, I call this the premature assumption.

If the premature assumption were right, then two other puzzles kick in. First, if children develop second-order attributions around six or seven and have the enabling resources they need, why do they wait several more years to do third- and higher-order attributions and continue to have problems with complex interpretive practices? In other words, why such gaps from five to seven for second-order attributions and then from seven to nine or even later for still higher-order attributions and complex interpretive practices, when all the elements required by the premature assumption are in place? And second, why is it that intelligent autistic people who pass the false-belief test, and possess the right enabling resources, nevertheless fail at higher-order attributions and complex interpretive practices (Baron-Cohen 1995; Baron-Cohen et al. 1993)?

My answer to these puzzles is to challenge the premature assumption on two fronts. The first is to separate metaintentional interpretation from propositional-attitude or metarepresentational interpretation and argue that the former does not and cannot afford higher-order attributions. This limitation

would explain why, even with the addition of enabling resources, higher-order attributions are not likely before the age of five. This is what I have endeavored to show in this paper. On the other front, the argument must show that only the metarepresentational categories of propositional attitudes afford higher-order attributions and that these categories develop after five, when a novel interpretive competence enters the scene. Aside from some hints planted earlier in this paper, this argument must await another occasion.

4. IF METAINTENTIONAL, THEN (MOSTLY) LEFT

There are neurological data that support the account sketched earlier. The primitive and modularized interpretive categories, such as those involved in face and gaze recognition, the representation of goal-directedness and agency, and recognition of emotions are located either in the left hemisphere or prefrontal cortex or amygdala (Baron-Cohen 1995; Baron-Cohen et al. 2000). These are the brain centers that govern the metaintentional interpretation of apes and very young human children. Damage to these centers can considerably delay or impair the capacity for interpretation in a cascade-type sequence, as in the case of autism. The left hemisphere is known to excel at selecting and processing a single, dominant mode of representation, and blocking out all the other. Regions of the prefrontal cortex are involved in executive functions, including the decoupling of current perceptual input from memory or imagination, the inhibition of the salient outputs of current perception and motivation. These are also the regions that contribute to the early representation of pretense and later the recognition of false belief. As noted earlier, the prefrontal cortex matures later than other brain centers and in particular matures its inhibitory capacity around the critical period of three to four.

By contrast, most of the metarepresentational work of later childhood

and adulthood involved in interpreting complex propositional attitude and in higher-order attributions is done mostly by the right hemisphere, often through sustained exchanges of information with the left hemisphere. The function of the right hemisphere is to form and network multiple, alternative and often conflicting sets of representations, particularly multiply embedded interpretive representations (Brownell et al. 2000; Frith and Frith 2000). Imaging studies of brain activity consistently show that first-order attributions light up the left and frontal part of the brain whereas second- and higher-order attributions light up the right side of the brain.

The last but not least important item of empirical evidence favoring the argument of this paper is clinical. Damage to the left and frontal brain, manifested in autism, compromises the ability to do first-order attributions. Autistic people fail to recognize emotions, have trouble with joint attention and usually fail false belief tests -- all metaintentional disabilities with an innate and probably genetic origin. By contrast, damage to the right hemisphere results in an inability to do higher-order attributions and to handle the complex interpretive practices based on such attributions. Such damage often produces schizophrenia, usually manifested late in adolescence or early youth, and the apparently resulting inability to do higher-order interpretation and engage in complex interpretive practices, including discourse processing, leaving intact the metaintentional capacity for first-order attributions (Corcoran 2000; Frith 1992; Frith and Frith 2000).

Brain imaging and clinical evidence thus suggest a localization contrast between the metaintentional left and partly frontal side of the interpretive brain and its metarepresentational right and partly frontal side. This contrast lends its support to the notion of two distinct interpretive competencies -- one metaintentional, older phylogenetically and operating earlier ontogenetically, the other metarepresentational, starting to operate at least two years after the false-belief revolution with new tools -- the category of propositional attitudes.

5. WIDER IMPORT

Although this paper proposes a fairly self-contained and domain-specific argument, it has a wider significance both in interpretation and beyond. In interpretation, the argument suggests that elaborate mental performances of apes and very young children in the areas of communication, socialization, pretense, imitation, strategic thinking involving two or more agents, and behavioral prediction are managed with resources that fall far short of metarepresenting attitudes, propositions and their higher-order concatenations. Some of these resources are nonrepresentational or even noncognitive and may have functions other than interpretive, yet they bring their decisive contribution to the formation or utilization of new interpretive resources (Bogdan 2000). In this sense, the human competence for interpretation is a hybrid construction and so are many of its specific abilities, including the category of (false) belief.

There are also implications beyond interpretation. I argued elsewhere that the competence for interpreting minds is involved in the acquisition or construction of novel mental faculties, such as understanding words, communicating linguistically, thinking reflexively and introspecting (Bogdan 2000; 2001). To this extent, the development and the level of sophistication of the interpretive competence can be viewed as a barometer of mental development in general. Viewed in this light, pre-inhibition and situated metaintentional interpretation indicates not just a mind confined to, and conscious only of, current perception and motivation, particularly when it comes to intentional relations. It also indicates a unirepresentational mind, so to speak, unable to entertain, maintain, network and go back and forth amongst several, often conflicting representations of such relations, either on- or off-line. It will take such a multirepresentational mind to do metarepresentational interpretation with propositional attitudes, often within the scope of other

attitudes. It is no accident, I think, that only such a mind can loop onto itself and become reflexive, introspective, self critical and exercising self control.

ACKNOWLEDGMENT

I am grateful to several audiences that heard and reacted to earlier fragments and versions of this paper read at the University of Arizona, University of Bucharest, the Belgian Society for Logic and Philosophy of Science, Université Catholique de Louvain, Institut Jean Nicod in Paris, and of course at the Wittgenstein symposium in Kirchberg. Special and warm thanks to Mircea Dumitru, Alvin Goldman, Josef Perner, Joelle Proust, Francois Recanati and Dan Sperber.

LITERATURE

Astington, J. and Gopnik, A. 1991, 'Developing Understanding of Desire and Intention', in A. Whiten (ed), Natural Theories of Mind, Blackwell, Oxford.

Baron-Cohen, S. 1995, Mindblindness, MIT Press, Cambridge.

Baron-Cohen, S. et al. (eds) 2000, Understanding Other Minds, 2nd edition, Oxford University Press, Oxford.

Bjorklund, D.F. 2000, Children's Thinking, Wadworth, Belmont.

Bogdan, R.J. 1997, Interpreting Minds, MIT Press, Cambridge.

Bogdan, R.J. 2000, Minding Minds, MIT Press, Cambridge.

Bogdan, R.J. 2001, 'Developing Mental Abilities by Representing Intentionality', Synthese, 129, 233-258.

Brownell, H. et al. 2000, 'Cerebral Lateralization and Theory of Mind', in S. Baron-Cohen et al. (eds). 2000, Understanding Other Minds, 2nd edition, Oxford University Press, Oxford.

Chandler, M. 1988, 'Doubt and Developing Theories of Mind', in J. Astington et al. (eds), Developing Theories of Mind, Cambridge University Press, Cambridge.

Cheney, D. and Seyfarth R. 1990, How Monkeys See the World, Oxford

University Press, Oxford.

Clements, W. and Perner, J. 1994, 'Implicit Understanding of Belief', Cognitive Development, 9, 377-397.

Corcoran, R. 2000, 'Theory of Mind in Other Clinical Conditions', in S. Baron-Cohen et al. (eds). 2000, Understanding Other Minds, 2nd edition, Oxford University Press, Oxford.

Flavell, J. et al. 1995, 'Young Children's Knowledge About Thinking', Monographs of the Society for Research in Child development, 60, 1.

Fodor, J. 1992, 'A Theory of the Child's Theory of Mind', Cognition, 44, 283-296.

Gopnik, A. and Astington, J.W. 1988, 'Children's Understanding of Representational Change' Child Development, 59, 26-37.

Leekam, S. 1991, 'Jokes and Lies: Children's Understanding of Intentional Falsehood', in A. Whiten (ed), Natural Theories of Mind, Blackwell, Oxford.

Leslie, A. 1987, 'Pretense and Representation', Psychological Review, 94, 412-426.

Leslie, A. 2000, 'Theory of Mind as a Mechanism of Selective Attention', in M. Gazzaniga (ed), The New Cognitive Neurosciences, MIT Press, Cambridge.

Lewis, C.: 1994, 'Episodes, Events and Narratives in the Child's Understanding of Mind', in C. Lewis and P. Mitchell (eds), Children's Early Understanding of Mind, Erlbaum, Hillsdale.

Meltzoff, A. 1995, 'Understanding the Intentions of Others', Developmental Psychology, 31, 838-850.

Perner, J. 1988, 'Higher-Order Beliefs and Intentions in Children's Understanding of Social Interactions', in J. Astington et al. (eds), Developing Theories of Mind, Cambridge University Press, Cambridge.

Perner, J. 1991, Understanding the Representational Minds, MIT Press, Cambridge.

Perner, J. and Lang, B. 2000, 'Theory of Mind and Executive Function', in

S. Baron-Cohen et al. (eds), Understanding Other Minds, 2nd edition, Oxford University Press, Oxford.

Russell, J., 1996, Agency, Erlbaum, Hillsdale.

Scholl, B. and Leslie, A. 2001, Minds, Modules and Meta-Analysis, Child Development, 72, 696-701.

Sullivan, K., Zaitchik, D. and Tager-Flusberg, H. 1994, 'Preschoolers Can Attribute Second-Order Beliefs', Developmental Psychology, 30, 395-402.

Tomasello, M. 1999, The Cultural Origins of Human Cognition, Harvard University Press, Cambridge.

Wellman, H. 1990, The Child's Theory of Mind, MIT Press, Cambridge.

Whiten, A.(ed): 1991, Natural Theories of Mind, Blackwell, Oxford.

Whiten, A. and Byrne, R. 1988, 'The Manipulation of Attention in Primate Tactical Deception', in R. Byrne and A. Whiten (eds), Machiavellian Intelligence, Oxford University Press, Oxford.

Whiten, A. and Perner, J. 1991, 'Fundamental Issues in the Multidisciplinary Study of Mindreading', in A. Whiten (ed), Natural Theories of Mind, Blackwell, Oxford.

Department of Philosophy

Tulane University

New Orleans, LA 70118

USA

and

Institut Supérieur de Philosophie

Université Catholique de Louvain

B-1348 Louvain-la-Neuve
Belgium