Why Self-Ascriptions Are Difficult and Develop Late

RADU J. BOGDAN
Tulane University

Many philosophers and a few psychologists think that we understand our own minds before we understand those of others. Most developmental psychologists think that children understand their own minds at about the same time they understand other minds, by using the same cognitive abilities. I disagree with both views. I think that children understand other minds before they understand their own. Their self-understanding depends on some cognitive abilities that develop later than, and independently of, the abilities involved in understanding other minds. This is the general theme of this chapter.

The argument focuses on what I take to be the core of understanding minds, namely, ascriptions of representational states or attitudes, such as desires or beliefs, whose relation (or directedness) to what they represent is registered in some fashion. I will dub this relation representingness (and treat it as equivalent to what philosophers call intentionality or aboutness.) The argument does not apply to nonrepresentational states, such as feelings, or states whose representigness is not at stake (as when I notice that I begin to see in the dark) or is not registered at all. To save space, I will shorten representational attitudes to (simply) "attitudes" and ascriptions of attitudes to (simply) "ascriptions." I call the ascriptions of one's own attitudes self-ascriptions and those aimed at the attitudes of others other-ascriptions. In what follows I will be concerned only with ascriptions that are sensitive to the representingness of the attitudes involved.

The argument is that self-ascriptions that register their own representingness are different, harder and later than other-ascriptions because the former require abilities that (a) are more complex than those required by the latter and (b) develop after the age of 4, when basic other-ascriptions are already in place. Section 1 documents these claims. Section 2 explains in general terms why self-ascriptions are harder than other-ascriptions. Section 3 argues that current theories fail to account for these

asymmetries. Section 4 turns to autobiographical memory for clues to the nature of self-ascriptions. Section 5 sketches a neuropsychological account of self-regulation and self-metarepresentation that explains why self-ascriptions develop later than other-ascriptions.

1. Asymmetries

The data presented below are rather sketchy and not uncontested but, if read as I suggest, they seem to favor the developmental and cognitive asymmetry I am proposing. First, there is some direct evidence of a developmental asymmetry. Children understand the desires of others around the age of two (Wellman 1990) but do not grasp their past unfulfilled or changed desires even a year later (Gopnik 1993). The same seems to be true of false belief: estimates by Astington and Gopnik (1988) place the self-ascription of false belief in the 4 to 5 interval but its other-ascription in the 3 to 4 year old interval (see also Flavell et al. 1995, 53-54, for a survey).

The indirect evidence is more robust and points in the same direction. The child's metacognition, required for an explicit awareness of self-attitudes, develops after 5 and takes time to mature; introspection is estimated to emerge several years later (Flavell et al. 1995; Nelson 1996). Before 7 to 8 children often fail to identify their own past thoughts and their contents, even when recent (Flavell et al. 1995, 80-81). If young children master self-ascriptions as soon as they master other-ascriptions and if, as seems likely, they master the basics of the latter around 4, why these delays in metacognition, introspection and thought identification? Couldn't it be because self-ascriptions develop late and slowly? Self-ascriptions, particularly of past attitudes, often require the inhibition of one's current cognition. Inhibition develops only after the 3 to 4 period (Bjorklund 2000; Houdé 1995; Leslie 2000).

There is also neurological evidence for asymmetry. The early theory of mind involved in face and gaze recognition and the representation of agency is done mostly in the left hemisphere (Baron-Cohen 1995). The left hemisphere excels at selecting and processing a single, dominant mode of representation, and blocking out all the others -- which is how the other-ascriptions of the first three years work. In contrast, most of the later

theory-of-mind work, including self-ascriptions, is done by the right hemisphere, in exchanges of information with the left hemisphere. The clinical evidence lends its support, too. Damage to the left hemisphere, manifested in autism, compromises the ability to handle joint attention and other-ascriptions of false belief -- all failures of early theory of mind. In contrast, damage to the right hemisphere -- in schizophrenia -- prevents complex self- and other-ascriptions, while leaving intact simple other-ascriptions (Corcoran 2000; Frith 1992).

A most telling evidence for the asymmetry thesis is that until 4 to 5, children lack autobiographical memory (Conway 2002; Nelson 1996). This sort of memory, which develops gradually, is needed in self-ascriptions of past attitudes. As far as I can tell, most experiments with and analyses of self-ascriptions of past desires and beliefs are about an immediate past (several minutes or hours). What the child has to do (which is not easy before 4) is inhibit the content of a current desire or belief and recall a past content. This is not enough for testing self-ascriptions that are sensitive to their representational relations. The memory of a past content is not the memory of a representational relation to that content.

It is the representational relations or representingness of one's own attitudes -- which I will call self-representingness -- that poses the mightiest challenge to self-ascriptions, delaying and lengthening its development, and making it harder to acquire and use than other-ascriptions. The notion of a sense of self-representingness is broad and allows for various resources -- procedures, images, schemes, thoughts -- that can do the job. The next section ventures a first explanation of why self-representingness is hard to register.

2. The Elusiveness of Self-Representingness

It helps to begin by making explicit what it takes to self-ascribe an attitude. I present only three conditions that matter to our discussion:

1 [evidential basis] having evidence for self-ascribing an attitude, in the form of inner experiences, mental activity, introspection, inference, etc.

2 [right concepts] possessing the appropriate concepts of attitudes,

such as desire or belief

3 [a sense of mental self-representingness] understanding one's own mental states as representational, as being about something (there is also a sense of bodily and behavioral self-directedness to a target)

The literature on early theory of mind often treats inner experiences of mental states and their contents as sufficient for self-ascriptions. This is not enough. Consider one's own desire. Although a desire is directed at something, to recognize having the desire, through some inner experience, is not to recognize its representingness. A child may recognize her desire from how it is experienced inside and from its content (what is desired) without recognizing the relation between experience and content. Many observations and experiments misdiagnose self-ascriptions by measuring only the inner experience of an attitude and some experience-based concept but missing the representingness of the attitude. Such misdiagnoses may actually suggest that young children manage self-ascriptions at the same age as, or even earlier than, they manage other-ascriptions.

Here is an analogy that underscores the importance of having a sense of representingness. Many organisms learn words or signs by associating a visual or sound experience with an object or event (as content or referent) without recognizing the referential relation between the experience and its content. This is the key difference between a parrot and a human child. The latter (but not the former) recognizes the relation in question because she is a naive psychologist who can grasp the reference relations between other minds and objects and events (Tomasello 1999). If this recognition is not factored into the analysis, it looks like word acquisition is a widespread animal trait -- just as self-ascription looks easy without the recognition of self-representingness. An experience of a mental state merely associated with its content is not good enough for the mastery of word meanings or self-ascriptions.

I will center my analysis on the self-ascriptions of past false belief. It is important to focus on self-ascriptions of past and mismatched attitudes, such as unfulfilled desires or false beliefs, rather than on self-ascriptions of current and successful attitudes, because only awareness of mismatch measures one's sense of self-representingness. One knows that one

mentally aimed at something when one knows that one missed it. This is why false belief tests the possession of the concept of belief. The pastness of an attitude forces a mental reconstruction that separates self-representingness from a current experience of an attitude or its current content. The self-ascription of a past false belief illustrates this point best because unlike one's own past desires or emotions, one's own past beliefs are more abstract and carry fewer experiential traces, so to speak. Whereas one can almost relive past desires or past emotions, one has to think (often hard) about past false beliefs.

In the false-belief experiments, the child needs to do at least the following to figure out the false belief of another person, whom I call target:

[A1] inhibit her propensity to treat what she currently knows about the facts of the case as the target's actual belief; and

[B1] recover a (visual) memory of the target's past relation to a content and represent that relation as the target's actual (and false) belief

In the experiments on self-ascription (such as the Smarties box that turns out to contain pencils), to figure out her own past false belief, the child must

[A2] inhibit her propensity to treat her current representation of the content (pencils) as her earlier belief; and

[B2] recover a memory of an earlier representation of the content (Smarties) and represent it as her earlier (and false) belief

There is an executive symmetry here: the first conditions, [A1] and [A2], call for inhibition -- a crucial development around 3 to 4, widely credited with enabling the ascriptions of false belief (Harris 1992; Leslie 2000; Perner 1991; also Bogdan 2003). The asymmetry between the two kinds of ascriptions is cognitive and concerns the representingness of the attitudes ascribed. Think of it intuitively: The child has no perceptual clues to the representingness of her own beliefs, that is, to the fact that her beliefs are representing (the relation) their contents; but she has such clues to the representational relation of a third-person belief. In [A1] the child sees not

only what the target perceives and therefore believes but also sees that the target has a perceptual and hence belief relation to the world. (As noted below, the young child's concept of belief seems to be modeled on that of perception.) But in [A2] all the self-ascriber perceives are pencils (the content); she does not perceive the relation to that content. In [B1] the child can recall visual memories about a past attitude of the target in relation to the content, whereas in [B2] the self-ascriber cannot recall any clues pointing to her belief as a relational attitude; all she remembers is a content -- the Smarties box. The representingness of her belief is not evident in her memory. And it could not be, because, before age 4, the child does not yet have an autobiographical memory that could retrieve her own past attitudes as representational relations to contents. The memories of early years are exclusively semantic and episodic (Conway 2002; Nelson 1996) and they deliver only contents experienced in some form. I develop these points in section 4. Right now I want to draw an implication from what has been said so far.

We should not underestimate the role of visual evidence in the development of ascriptions. Several decades of research have shown that the child's (and possibly the ape's) understanding of other minds begins and develops through visual observation of others -- their faces and facial expressions, looks, gaze, bodily posture, movement, and behavior -- and of the contexts of social interaction. Congenitally blind children take longer to develop an understanding of other minds (Hobson 1993). The more complex and invisible attitudes, such as belief and intention, may be modeled on the simpler and more visible ones, such as perception and desire (Wellman 1990). It is likely that the early sense that children have of the representingness of other-attitudes emerges out of their visual recognition of gaze and attention (Tomasello 1999; Moore and Dunham 1993). The point I am making is not that a theory of mind develops out of the visual observation of others. Many species engage in such observation but very few do theory of mind. The point, rather, is that visual evidence is important for activating, maturing and guiding the theory-of-mind abilities to detect other-attitudes.

There is no corresponding visual evidence for detecting self-representingness. Children represent the contents of their own attitudes and sense their presence in the mind from internal signals but cannot

observe their representingness as relations to contents. People do not observe their perception or belief as relations to the world, particularly in the case of past such relations. Yet children develop nonvisual means of registering self-representingness from inside their minds. How do they do it? A brief survey of current theories does not yield a satisfactory answer to this question.

3. Other Views

The main accounts of self-ascriptions are the theory-theory view and the simulation view. I locate my own view relative to them by reversing the predictions of failure that each makes against the other. The theory-theory predicts that simulation theory, by positing self-ascriptions to be prior to and easier than other-ascriptions, would be distressed by evidence of temporal and cognitive symmetries between the two sorts of ascriptions. Simulation theory predicts that theory-theory, by positing such symmetries, would be distressed by evidence of asymmetries favoring self-ascriptions as earlier and easier. Against the theory-theory, my analysis proposes a temporal and cognitive asymmetry, but, against simulation theory, it finds other-ascriptions earlier and easier than self-ascriptions.

The theory-theory view is committed to the developmental and conceptual symmetry between other- and self-ascriptions (Gopnik 1993; Gopnik and Meltzoff 1997; Perner 1991). I take the general case made for asymmetry in section 1 to count against the theory-theory position. But I also have more specific arguments. The theory-theorist may agree that there are differences between self- and other-ascriptions with respect to evidence and cognitive resources, such as inhibition and memory, but insists that they need not entail a difference in the concepts utilized (Gopnik 1993; Gopnik and Wellman 1992; Malle, personal communication; Perner 1991). And sameness of concepts appears to invalidate the asymmetry thesis. Actually, it doesn't, as I argue next.

First, my asymmetry hypothesis is that, helped by visual evidence, young children first apply the concepts of attitudes to others, and need more time, more cognitive effort, different evidence, and new abilities to figure out how to apply the (alleged) same concepts to themselves. Same

concepts notwithstanding, it is still the case that self-ascriptions can be harder and develop later than other-ascriptions. A second argument points to evolution. The asymmetry is not just a matter of evidence and cognitive resources. It is also a matter of the function of a theory of mind. A good case can be made that in humans and possibly other primates a naive theory of mind first evolved to deal with conspecifics, not with selves. The most pressing challenges posed by conspecifics are in ongoing contexts of social interaction and involve observable features of conspecifics, such as facial expression, gaze, bodily posture, behavior, communication, and the like (Bogdan 1997; Tomasello 1999; Whiten 1991). It is unlikely that the basic concepts and ascription strategies of theory of mind evolved for reasons other than registering visible relations between conspecifics and the world, ascriber and conspecifics, and among conspecifics themselves. Looked at in this light, the turn to self-ascription is a late evolutionary and developmental event, and not a companion counterpart to other-ascriptions.

The argument from evolution and the ones that follow begin to challenge the same-concepts claim. Suppose that young children are primarily interested in other people and that they first ascribe to them observable attitudes, such as seeing. Suppose also that these ascriptions are directed at concrete, spatio-temporally defined items in the world, such as objects and events, and not at propositions, as they will be later. Suppose, finally, that these early other-ascriptions are egocentric, in that they reflect the child's ongoing motivation and perception, and situated because tied to current contexts of interaction with others. I have defended this characterization of early theory of mind elsewhere (Bogdan 1997; 2000; 2003), so I will move to the relevant arguments it entails.

One argument is this: Theory-theorists tend to think that (unless innate or essentialist) concepts are normally formed and revised in response to perceptually accessible facts and regularities in their domains (Gopnik 1993; Gopnik and Meltzoff 1997; Perner 1991; Wellman 1990). If that is so, then it is hard to see how other-concepts could initially be the same as self-concepts, when the facts, regularities and the perceptual evidence revealing them in the other-domain are so different from those in the self-domain in key respects, particularly concerning self-representingness. This initial difference does not preclude a later alignment of self- to other-concepts,

when the older child acquires a sense of her own mental representingness.

Another argument is the following: Theory-theorists agree that early theory of mind is not metarepresentational, for it does not represent the rather abstract and invisible representational relations of complex attitudes, such as opinions, intentions, or hopes. However, one's own past attitudes, particularly false or unfulfilled ones, also have a rather abstract and invisible representingness. The early concepts for other-ascriptions are not equipped to handle this sort of past self-representingness -- and I do not think they do. After the metarepresentational turn around age 4, the child acquires a new supply of concepts that track self-representingness. The question is whether the new concepts are self-other symmetrical or not. It is a question I will tackle in section 5. The answer is less simple than expected.

I turn next to simulation view, with its two versions. The practical-reasoning version of Robert Gordon (1986), which does not require ascriptional concepts, has two options. One is an ascent routine that habituates the child to the link between a content in mind and first-person locutions, such as 'I believe that ...' I do not see how this works for 'I falsely believe that...' and particularly for 'I believed falsely that...' without begging the question at stake, which is how one's own past representingness is represented. As Gordon (1993, 45) notes, the ascent routine cannot deliver it. The routine provides at best a head start for counterfactual and imaginative simulation, which is the other option. How would this option work?

According to Paul Harris (1992), in the test of one's own past false belief, the child must imagine the proposition she originally entertained and took to be true (smarties in the box) and, inhibiting her current knowledge (of pencils), report on its usual contents (smarties). Before 4, children have the same difficulty representing their own past false beliefs as well as those of others, for they do not have enough counterfactual imagination. But how would that imagination track self-representingness? Young children imagine mostly in visual terms and those, I argued earlier, are not the right terms for one's own past representingness.

How about introspective simulation based on ascriptional concepts (Harris 1992; Goldman 1993)? This view advocates an asymmetry in the opposite direction from mine: self-ascriptions develop earlier and are easier than other-ascriptions. One problem is that the introspection that classifies

attitudes need not reach beyond experience, current or recalled. As Alvin Goldman notes (1993, 105), introspection can identify the type and content of an attitude but not its representingness and hence truth value. Another problem is that a representingness-sensitive introspection may be a late development, perhaps as late as age 7 or 8 (Flavell et al. 1995), which is later than the onset of self-ascriptions. Like Harris, Goldman (1993, 43) thinks that the key obstacle for a young introspector attempting to recognize her own past false belief is sorting out and dating conflicting representations. This is right, but even with representations of current and past contents sorted-out and dated, the young introspector still has no idea of her former self being representationally related to and actually misrepresenting a past content. She just recalls a past content that is different from a current one.

Theory-theory and simulation are not the only accounts of self-ascriptions. There are modular accounts, such as those of Leslie (2000) and Baron-Cohen (1995), and also accounts of forms of self-other coordination, such as those of Gopnik and Meltzoff (1997) and Barresi and Moore (1996). And the list is not over. I do not have the space to discuss them but I will say this much, without argument: As survival devices, modules evolved to deal with other-attitudes, not one's own. Their sensitivity to faces, eye, gaze, and bodily signals is evidence of sensitivity to others and their representational relations to the world. Nichols and Stich (2003) posit an innate self-monitoring module, active since age 2, which detects one's own mental states. The problem is that detection (like inner experience and introspection) may signal the presence and partly the type of an attitude but not its representational relation. Coordination mechanisms map one's experiences onto the attitudes of others, not of selves, and such mappings need not be sensitive to representational relations. In general, modular and coordination accounts best explain the other-ascriptions of early childhood but not later self-ascriptions (Bogdan 1997; 2000).

Brief and fast-paced as my critical survey has been, it should be recalled that it concerns solely views about self-ascriptions that are sensitive to the representational relations of attitudes (the hard ones) and not other sorts of self-ascriptions -- say, of feelings or of current or immediately past mental states recognized solely through their internal

experiences or their contents (the easy ones). Such experiences and the partial concepts based on them are necessary, as early precursors, for the development of self-ascriptions -- a necessity explained in various ways by the views just surveyed. But their story remains incomplete because a sense of self-representingness is not in the picture.

If the early theory of mind is insensitive to self-representingness, what are the resources that generate that sensitivity and why do they develop only after 4? I propose to look for clues to an answer in the domain of memory, because it plays a major role in past self-ascriptions and because its development parallels in important respects that of theory of mind.

4. Memories of Past Attitudes

We have immediate access only to our current attitudes. Past attitudes must be retrieved from memory. The question is what sort of memory could do this job. Young children have good memory for things, events and situations. This is semantic memory. It retrieves only content, without its actual experience. Another sort of memory, called episodic, retrieves experiential details of past contents, linked to spatio-temporal contexts, and vivid reactions and emotions. Episodic memory also operates in young children and perhaps some nonhuman species (Clayton 2002). Episodic memories are represented in the same brain areas as are actual experiences (Conway 2002). This is significant because it suggests that, like ongoing mentation, episodic memory accesses only past experiences or their contents but not past self-representingness. Furthermore, the reliving of past experiences or contents may recreate a vivid sense of having seen or desired something but not of having believed something, which is a much less vivid attitude. This is why one's own past perceptions or desires are easier to recall than past beliefs.

An objection raised by Bertram Malle (personal communication) touches on both episodic memory and the central issue of this paper. Why should a self-ascriber need a separate sense of self-representingness? Wouldn't it be enough to re-experience episodically one's desire (say) for coffee this morning, which may combine memories tagged with the time it

happened, who the self was, what was desired, and the sort of mental state it was? And doesn't one do the same with self-ascriptions of a current attitude -- join an experience of the type of attitude with an experience of its content and of the self that has these experiences? Aren't young children capable of all these exploits, both in episodic memory and current self-ascriptions?

Suppose they are. It is still a developmental fact that children recall things episodically very early but cannot do self-ascriptions that are sensitive to their representingness until a few years later. It is also a developmental fact, proven by infantile amnesia, that the young children's episodic memories evoke relatively short-lived experiences, and that access to such memories tends to degrade rather quickly (Conway 2002). The reason, I think, most experiments with young children's self-ascriptions appeal only to a recent past is that they test only their episodic memory. If older children and adults have long-term memories, it can't be solely because they have episodic memory. I suspect that long-term memories have something to do with being able to represent one's past attitudes and their representingness. So what sort of memory could do it?

The answer is autobiographical memory. It is the sort of memory that children lack until around 3 and half to 4, and that develops gradually until 6 (Nelson 1996, 157 and 162; Bjorklund 2000, 264), which is also the interval when representingness-sensitive self-ascriptions develop. Autobiographical memory terminates infantile amnesia by integrating and consolidating episodic memories in autobiographical terms and enabling a retrieval of past attitudes. How does this work? Autobiographical memory is said to add to its episodic basis an autonoetic or quasi-introspective consciousness and recreative thinking. These new abilities seem uniquely human and develop in 4 to 6 interval (Conway 2002). I schematize the ontogenesis of autobiographical memory as follows:

> semantic memory + [recreation of experiences in terms of perceptual vividness, spatio-temporal framing, and affective associations] = episodic memory

episodic memory + [autonoetic consciousness + recreative thinking] = autobiographical memory?

Yet reliving past experiences consciously and recreatively may still not be enough for autobiographical memory. One may recall episodically, through imagery or inference, the passing show of past events without representing one's past attitudes as true or false or referring to this or that. What else is needed? Josef Perner's answer (1991, 163-169; 2000) is metarepresentation. It explains why episodic memory turns autobiographical and why (on my analysis) the latter can represent one's past self in various representational relations to things and situations. *This idea is plausible for two reasons. In remembering autobiographically past attitudes, we represent ourselves back then representing the contents of past attitudes. This is what metarepresentation does. The idea is also plausible because, on most accounts, metarepresentation develops around 4. The sort of metarepresentation Perner has in mind is symmetrically shared by self- and other-ascriptions. For reasons discussed earlier, it does not seem quite the right sort. If autobiographical memory is required for self-ascriptions of past attitudes, we need to look for another sort of metarepresentation, one that is intrinsically sensitive to self, as is autobiographical memory itself. This is the issue I turn to next and last.

5. Minding Our Own Minds

My proposal is that a sense of self-representingness grows out of the executive tasks of self-regulation of the new mental activities that develop after the age of 4. The main self-regulatory tasks consist in of holding in mind many representations in an active state for an extended period, monitoring, controlling, integrating and manipulating the information needed for a task, and inhibiting task-irrelevant information. Most of this work can only be done in terms of what and how one's thoughts represent what they do. This is why this intramental work calls for a sense of one's own thoughts being related to what they represent -- a sense of mental self-representingness. The demonstration of this thesis can only be sketched here in a few telegraphic steps.

I begin by contrasting two metaphors. Until the 3-4 interval the young mind operates on a single central screen, where perceptual and memory inputs are displayed and constantly updated by new inputs. It is a mind largely, though not entirely, confined to current motivation and perception. The young mind can imagine beyond the current inputs but still within their frame and theme. Think of the imaginative stance of young childhood as a sort of little screen or box that opens in a corner of, and from inside, the larger screen dominated by current perception and/or memory. This is a simplification, of course, but the contrast it highlights is real.

After 3, the young mind is shaken by several mental commotions, executive as well as cognitive, and revolutionary in their cumulative impact. Chief among them are the inhibition of current perception, the linguistic recoding and representational explicitation of earlier procedural competencies, the development of short-term memory as the workspace where multiple and alternative representations can be maintained, manipulated and integrated in various formats (Diamond 2001; Houdé 1995; Karmiloff-Smith 1992). These developments liberate the young mind from the captivity of single-screen or uniplex mentation and enable it to entertain simultaneously, in different but interconnected mental screens, nested sets of alternative and often conflicting representations of actual and nonactual, current, past and counterfactual situations. A multi-screen or multiplex mentation comes of age. It creates its own pressures for internal self-regulation in the form of supervisory capacities operating explicitly on and with thoughts in terms of their representational relations and features (Perner 1998; Shallice and Burgess 1993). The child's mind thus develops an internal metarepresenter or metamind. The chief neural platform of this new metamind is the (dorsolateral) prefrontal cortex and the integrative connectivity handled mostly by its right hemisphere and reaching across large regions of the brain. The growth of this platform is most dramatic in the 3 to 6 interval (Diamond 2001).

Nothing in the story so far mentions theory of mind. The self-regulatory job of the metamind is the basic phenomenon. It is quite a different question whether, in order to carry out its self-regulatory functions, the metamind develops its own metarepresentational tools or recruits those of the child's theory of mind. Similar or nearby brain

structures, which develop in the 4 to 7 interval, seem to have a hand in many executive and theory-of-mind tasks. Such neural and temporal proximity and the idea of control by metarepresentation may suggest that it is developments in theory of mind that are co-opted for intramental self-regulation (Perner 1998; to some extent Frith 1992).

This scenario is possible but unlikely, I think, because of the very nature of what is to be monitored and controlled, and how. Consider intention. It is hard to see how metarepresenting one's own intentions, for self-regulation, could result simply from recruiting a theory-of-mind concept. Intentions are recognized and monitored internally in order (among other things) to distinguish between actions caused by our desires and plans and those reacting to external events. Failure to make this distinction impairs the act of intending and may result in delusions of control and other passivity experiences (Frith 1992). Failure to make the same distinction in the case of other people is unlikely to have similar effects. One must first have an internal sense of what it is to monitor and control one's own acts and representations, whether sensorimotor or mental, before one can conceptualize such acts and representations.

There is also a neurological reason for doubting that developments in theory of mind are responsible for self-metarepresentation. Most of the latter work is done by the right hemisphere in exchanges with the left hemisphere, whereas early other-ascriptions activate mostly the task- and domain-specific left hemisphere (Brownell et al. 2000). And, as noted in section 1, damage to the left and frontal brain affects other-ascriptions, whereas damage to the right hemisphere impairs only higher-order ascriptions and self-metarepresentation (Corcoran 2000).

Fortified by these reasons, I propose that in the 4 to 7 interval between age 4 and 7, the self-regulatory metamind develops its own predispositions for self-metarepresentation and thus triggers the development of a sense of self-representingness.

To get a better handle on this proposal and see its cerebral plausibility, consider the following (much simplified) analogy. To monitor and control its motor actions, an organism must have metamotor information. Suppose its actions are represented by motor images that track bodily positions relative to visual stimuli from the targets of its actions. If the organism is just

reactive or on automatic pilot, the motor images are fed into preset action schemas, and all is well. But if the organism is endowed with top-down control and attention, and needs to initiate a new action or modify an action or watch closely what it is doing, it must be able to track the motor images themselves as they relate to their targets. *Tracking the representational relations of motor images is the job of metamotor images as second-order motor representations. Whereas first-order motor images represent actions relative to bodily states and external targets, the metamotor images compare what first-order motor images represent with internal models from motor memory and with action predictions made by a planning center (Damasio 1999; Jeannerod 1997). The metamotor comparisons between first-order motor images and memories and planning predictions enable an organism to register and control the representational relations of its motor images because the results of the comparisons provide information about the organism being related to a target and about whether it is on target, properly directed at it or not, and if not, by how much. Metamotor images provide a sense of motor self-representingness because they enable an organism to do things with and to motor images in terms of their representational relations.

I suggest we think in the same control-of-activity spirit about the self-regulatory work of the metamind. Multiplex mentation is a new domain of activity to be mapped and supervised. It happens to be an intramental domain, inhabited by one's own thoughts and thought processes. Given that thoughts represent all sorts of targets (worldly, mental, abstract), and cause as well as get feedback from other thoughts in terms of what they represent, the self-regulatory work of the metamind must be metarepresentational and engage thoughts at their representational joints, such as reference, coherence, and truth value. As a result, the metamind generates a sense of one's own thoughts being related to what they represent, which is a mental sense of self-representingness.

So construed, self-metarepresentation may have a generic format that treats one's own thoughts as mental states that represent and have internally recognizable functions (to remember, to infer, to act on, etc.) but are not yet classified as desires, beliefs, intentions, and so on, according to

a theory of mind. It is when one's metamind must represent one's own thoughts as theory-of-mind attitudes that this generic format may become explicitly structured by theory-of-mind concepts. Some of these concepts probably build on their precursors, particular on the internal symptoms of desires, beliefs, and so on. It is equally possible that the executive demands on the emerging metamind may force a dramatic revision in the child's earlier theory of mind, if the latter is to integrate self- and other-ascriptions in ways that have self-regulatory impact. After all, developments after 3 to 4 acquaint and confront the child's mind with an entirely new domain -- her own thoughts now recognized explicitly as representational. Her theory of mind must adapt to this new domain and reconcile it with the domain of other minds. An outcome of this process may be a new conceptual cartography of the mind, integrating self- and other-ascriptional concepts. But the point I am emphasizing here is that the initial self-metarepresentation is likely to result from intramental self-regulation.

To return to our parallel topic for a last show of support, an internally driven and attitude-free self-metarepresentation seems also involved in autobiographical memory. One's episodic memories beam back vivid snippets of an original experience, testifying to its authenticity. What confers a sense of self-representingness and veracity to those memories is the monitoring, integration and evaluation of the representations involved in terms of how they fit together, how they organize the information thematically and narratively, and so on. Autobiographical memories convey a sense of their representingness without necessarily representing self-attitudes in a theory-of-mind format. One need not have a past belief about an event in order to remember the event autobiographically. It is the other way around. One remembers autobiographically the event because of how the work of one's memory meshes with that of one's metamind. The resulting memories in turn allow the recovery of past attitudes toward the remembered event.

Time to sum up. If we ask why self-ascriptions are later and harder than other-ascriptions, the answer I have proposed is that, unlike the latter, the former are grounded in an internally driven self-metarepresentation. Whereas many of the concepts and schemes of other-ascriptions emerge early in the child's theory of mind, the self-metarepresentation required for

self-ascriptions develops only after the age of 4, for neuropsychological reasons having to do with brain development and self-regulation rather than theory-of-mind.

ACKNOWLEDGMENTS

My warm thanks to Janet Astington, Sara Hodges, Carolyn Morillo, and Josef Perner for helpful comments; to audiences at the University of Oregon, Universidad Autonoma de Barcelona, and University of Bucharest for their reactions to early versions of this paper; and particularly to Bertram Malle for detailed observations and excellent suggestions.

REFERENCES

Astington, J. and Gopnik, A. (1991). Developing understanding of desire and intention. In A. Whiten (ed), Natural theories of mind. Oxford: Blackwell.

Baron-Cohen, S. (1995). Mindblindness. Cambridge: MIT Press.

Barresi, J. and Moore, C. (1996). Intentional relations and social understanding. Behavioral and Brain Sciences 19, 107-122.

Bjorklund, D.F. (2000). Children's thinking. Belmont: Wadworth.

Bogdan, R. (1997). Interpreting minds. Cambridge: MIT Press.

Bogdan, R. (2000). Minding minds. Cambridge: MIT Press.

Bogdan, R. (2003). Watch your metastep: The first-order limits of early intentional attributions. In C. Kanzian et al (eds), Persons. Vienna: obv&hpt.

Brownell, H. et al. (2000). Cerebral lateralization and theory of mind. In S. Baron-Cohen et al. (eds). Understanding other minds, 2nd edition. Oxford: Oxford University Press.

Clayton, N. et al. (2002). Elements of episodic-like memory in animals. In A. Baddeley et al (eds), Episodic memory, Oxford: Oxford University Press.

Conway, M. (2002). Sensory-perceptual memory and its context: autobiographical memory. In A. Baddeley et al (eds), Episodic Memory. Oxford: University Press.

Corcoran, R. (2000). Theory of mind in other clinical conditions. In S. Baron-Cohen et al. (eds). Understanding other minds, 2nd edition. Oxford:

Oxford University Press.

Damasio, A. (1999). The Feeling of what happens. New York: Harcourt.

Diamond, A. (2001). Normal developments of prefrontal cortex from birth to young adulthood. In D.T. Stuss and R.T. Knight (eds), The Frontal lobes. Oxford; Oxford University Press.

Flavell, J. et al. (1995). Young children's knowledge about thinking, Monographs of the Society for Research in Child development, 60, 1.

Frith, C. (1992). The neurological basis of schizophrenia. Hillsdale: Erlbaum.

Goldman, A. (1993). The psychology of folk psychology. Behavioral and Brain Sciences, 16, 15-28.

Gopnik, A. (1993). How we know our minds. Behavioral and Brain Sciences, 16, 1-14.

Gopnik, A. and Astington, J. (1988). Children's understanding of representational change. Child Development, 59, 26-37.

Gopnik, A. and Wellman, H. (1992). Why the child's theory of mind really is a theory. Mind and Language, 7, 145-171.

Gopnik. A, and Meltzoff, A. (1997). Words, thoughts and theories. Cambridge: MIT Press.

Gordon, R. (1986). Folk psychology as simulation. Mind and Language, 1, 158-171.

Gordon, R. (1993). Self-ascriptions of belief and desire. Behavioral and Brain Sciences, 16, 45-46.

Harris, P. (1992). From simulation to folk psychology. Mind and Language, 7, 120-144.

Hobson, R.P. (1993). Autism and the development of mind. Hillsdale: Erlbaum.

Houdé , O. (1995). Rationalité , dé veloppement et inhibition. Paris: Presses Universitaires de France.

Jeannerod, M. (1997). The cognitive neuroscience of action. Oxford: Blackwell.

Karmiloff-Smith, A. (1992). Beyond modularity. Cambridge: MIT Press.

Leslie, A. (2000). Theory of mind as a mechanism of selective

attention. In M. Gazzaniga (ed), The new cognitive neurosciences. Cambride: MIT Press.

C. Moore and P.J. Dunham (eds) (1995). Joint attention. Hillsdale: Erlbaum

Nelson, K. (1996). Language in cognitive development. Cambridge: Cambridge University Press.

Nichols, S. and Stich, S. (2003). Mindreading. Oxford: Oxford University Press.

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

Perner, J. (1998). The meta-intentional nature of executive functions and theory of mind. In P. Carruthers and J. Boucher (eds), Language and thought. Cambridge: Cambridge University Press.

Perner, J. (2000). Memory and theory of mind. In E. Tulving et al (eds), The Oxford handbook of memory. Oxford: Oxford University Press.

Shallice, T. and Burgess, P. (1993). Supervisory control of action and thought selection. In A. Baddeley and L. Weiskrantz (eds), Attention. Selection. Awareness and Control. Oxford: Oxford University Press.

Wellman, H. (1990). The Child's theory of mind. Cambridge: MIT Press.

Whiten, A. (ed.) (1991). Natural theories of mind. Oxford: Blackwell.