What Makes Human Cognition Unique? From Individual to Shared to Collective Intentionality

MICHAEL TOMASELLO AND HANNES RAKOCZY

Abstract: It is widely believed that what distinguishes the social cognition of humans from that of other animals is the belief-desire psychology of four-year-old children and adults (so-called theory of mind). We argue here that this is actually the second ontogenetic step in uniquely human social cognition. The first step is one year old children's understanding of persons as intentional agents, which enables skills of cultural learning and shared intentionality. This initial step is 'the real thing' in the sense that it enables young children to participate in cultural activities using shared, perspectival symbols with a conventional/normative/reflective dimension—for example, linguistic communication and pretend play—thus inaugurating children's understanding of things mental. Understanding beliefs and participating in collective intentionality at four years of age—enabling the comprehension of such things as money and marriage—results from several years of engagement with other persons in perspective-shifting and reflective discourse containing propositional attitude constructions.

By all appearances, the cognitive skills of human beings are very different from those of other animal species, including our nearest primate relatives. Human beings and only human beings cognize the world in ways leading to the creation and use of natural languages, complex tools and technologies, mathematical symbols, graphic symbols from maps to art, and complicated social institutions such as governments and religions. The puzzle is that other primates have created none of these things even though some—the great apes—are as closely related to humans as horses are to zebras, lions are to tigers, rats are to mice.

The solution to the puzzle is that such things as languages, symbolic mathematics, and complex social institutions are not individual inventions arising out of humans' extraordinary individual brainpower, but rather they are collective cultural products created by many different individuals and groups of individuals over historical time. And so if we imagine a human child born onto a desert island, somehow magically kept alive by itself until adulthood, it is possible that this adult's cognitive skills would not differ very much—perhaps a little, but not very much—from those of other great apes.

We would like to thank Tanya Behne, Malinda Carpenter, Wolgang Detel, Bob Gordon, Samuel Guttenplan, Ulf Liszkowski, Larry Roberts, and Sebastian Rödl for helpful feedback on the paper. A special acknowledgement goes to Heide Lohmann whose work we draw upon heavily in the second half of the paper, and whose ideas on language and theory of mind were instrumental in our thinking.

Address for correspondence: Max Planck Institute for Evolutionary Anthropology, Inselstrasse 22, D-04103 Leipzig, GERMANY
Email: tomas@eva.mpg.de

Mind & Language, Vol. 18 No. 2 April 2003, pp. 121–147.
© *Blackwell Publishing Ltd. 2003*, 9600 Garsington Road, Oxford, OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

UNCORRECTED PROOF

This person would certainly not invent by him or herself a natural language, or algebra or calculus, or science or government. And so perhaps it is the case that the uniquely human cognitive skills that make the most difference are those that enable individuals of the species *Homo sapiens* to, in a sense, pool their cognitive resources, that is, to create and participate in collective cultural activities and products. When viewed from the perspective of the individual mind, these cognitive skills of cultural creation and learning may not differ so very much from those of other primate species.

The most fundamental cognitive skills involved in processes of cultural creation and learning are those involved in the understanding of persons (sometimes called, misleadingly, 'theory of mind'). Thus, Tomasello, Kruger, and Ratner (1993) and Tomasello (1999) argued and presented evidence that a number of different forms of social and cultural interaction and learning depend fundamentally on the way human individuals understand one another. When one year old children understand adults' behavior as intentional and their perception as attentional (i.e., understand them as intentional agents[1]), they are able to interact with them and to learn from them in some unique ways. When four-year-olds understand that others have thoughts and beliefs that may differ from reality (i.e., understand them as mental agents), they are able to engage in still other types of social and cultural interactions and learning. Although a number of theorists have proposed that human beings engage in unique forms of social cognition, the proposal of Tomasello and colleagues is distinguished by its emphasis on the connection of these skills to culture and cultural learning, including language, and in its emphasis on the primacy of understanding persons as intentional agents for processes of human culture—with the understanding of persons as mental agents representing a kind of 'icing on the cake'.

It may still turn out that some nonhuman primates understand some aspects of the goal-directed actions of other individuals—a question we address specifically, albeit briefly, later. But our primary concern in this essay is how young children use an understanding of persons as intentional agents to participate in what are unarguably uniquely human forms of social and collective intentionality such as linguistic communication, shared pretense, and discourse about mental processes. In examining these phenomena, we make three basic claims.

- Human beings have a biological adaptation for a species-unique form of social cognition. This adaptation expresses itself ontogenetically at two key developmental moments, one at about one year of age and one at about four years of age. Although conceptualized and investigated in very different ways—as skills of joint attention and theory of mind, respectively—these are really just two phases of the same developmental pathway: understanding persons as intentional agents and then as mental agents.
- Understanding and coordinating with intentional agents at one year of age is the truly momentous leap in human social cognition in the sense that it already distinguishes human beings from other primates, and it enables

[1] 'Intentional' here is used in the sense of 'acting with an intention'.

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

human children to participate in and master cultural activities of all kinds, including linguistic communication. In participating in cultural activities, two year old children demonstrate their ability to establish self-other equivalence, to take different perspectives on things, and to reflect on and provide normative judgments of their own cognitive activities. We thus call these activities shared intentionality.

- Three and four year old children's coming to understand mental agents—who have thoughts and beliefs that may be false—depends both on the understanding of intentional agents and on a several year period of continuous interaction, especially linguistic interaction, with other persons. Based especially on their participation in perspective-shifting and reflective discourse, some new kinds of normativity emerge—specifically, those involving beliefs (with intensionality and norms of rational inference and truth), which in turn enable the comprehension of cultural institutions based on collective beliefs and practices such as money and marriage and government. We thus call these activities collective intentionality.

1. Understanding Intentional Agents

It is commonly believed that what most clearly distinguishes the social cognition of humans from that of other animals is the belief-desire psychology with which adult humans perceive and describe one another as practically and epistemically rational subjects. And, as usual, there are various proposals to the effect that this belief-desire psychology is an innate component of the human mind (e.g., Baron-Cohen, 1995; Leslie, 1994; Fodor, 1992). Following a long tradition in Western epistemology, the mental state of belief is given privileged status theoretically as the mark of the mental. Beliefs are fully mental because they are independent of reality in the sense that there can be false beliefs, so that, for example, the truth value of the proposition 'I believe that it is raining' is independent of the truth value of the embedded proposition 'It is raining'. In addition, the ability to understand beliefs is sometimes characterized as the ability to engage in meta-representation, and, relatedly, beliefs carry with them a normative quality insofar as they may be either true or false. Meta-representation and evaluation imply that subjects can take a reflective stance toward themselves and their own cognitive activities, observing and evaluating their interactions with the world. Young children are able to understand and reflect on false beliefs at around 4 to 5 years of age.

A number of proponents of this general view have also become interested in some 'precursors' of human belief-desire psychology, with the implication that these do not yet concern fully mental phenomena. These precursors involve young children's ability to understand and deal with simpler psychological states, such as the perceptions, intentions, attention, emotions, and desires of other people, and to interact with them in various kinds of joint attentional activities (e.g., Baron-Cohen, 1993; Wellman and Bartsch, 1994). These psychological states are not as clearly quarantined

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

from the real world as are beliefs, so that utterances like 'I see it is raining' (non-epistemic seeing) and 'I want to go there' are not referentially opaque and meta-representational in the same way as statements involving explicitly indicated beliefs[2]. These 'precursors' begin to emerge at around 1 to 2 years of age.

But there is another way to look at things. From an evolutionary point of view, what seems to distinguish the social cognition of humans from that of other animals is the ability to deal with any psychological states at all, including simpler mental states such as intentions and attention (Tomasello and Call, 1997; Povinelli, Bering and Giambrone, 2000). Moreover, understanding these simpler mental states would seem to be sufficient for young children to master the use of cultural artifacts and symbols of various sorts, including linguistic symbols, which they do from shortly after their first birthdays—and in virtually everyone's account, the ability to create and use linguistic symbols is a key distinguishing feature of human social cognition. Therefore, from an evolutionary point of view it might be more perspicacious to say that human beings, and only human beings, evolved the ability to understand and reason about the psychological states of persons. This ability first manifests itself in human ontogeny at around one year of age in the understanding of such things as intentions and attention, and it then develops further towards a full-fledged belief-desire psychology in the following few years.

Although this could be seen as nothing more than a rhetorical point—privileging the understanding of intentions over beliefs—it is actually a substantive proposal with empirical predictions. The proposal is that the key human biological adaptation was for understanding persons as intentional agents, and the understanding of persons as mental agents possessing beliefs is an ontogenetic construction that depends not only on this adaptation but on several years of certain kinds of social and linguistic interactions—with no specific biological underpinnings of its own. Key to such an account is to show that the understanding of intentional agents at one year of age is 'the real thing' in the sense that it concerns fully mental states and so has within it the seeds of the later-emerging and more powerful belief-desire psychology. As evidence for this view we document in what follows that one year old social cognition, and the joint attentional activities it enables, manifests three key characteristics:

(i) 'sharedness', involving self-other equivalence;
(ii) an understanding of perspective, involving the construal of the same thing under different descriptions;
(iii) an appreciation of normativity, involving a reflective stance.

The first two of these may be seen in one-year-olds' joint attentional activities and in their understanding of the intentions and attention of other persons, and these will be described in the immediately following sub-section. The third characteristic

[2] For some ways of specifying these kinds of differences see, e.g., Barwise and Perry (1983), Perner (1991).

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

is most readily apparent in one-year-olds' use of linguistic symbols and other cultural artifacts, and these will be described in the two following sub-sections.

1.1. The Nine-Month Revolution

Beginning before the beginning, we may observe that 6-month-old infants interact dyadically with objects, grasping and manipulating them, and they also interact dyadically with other people, expressing emotions back-and-forth in a turn-taking sequence. But at around 9–12 months of age a new set of behaviors begins to emerge that are triadic in the sense that they involve a referential triangle of child, adult, and the object/event to which they share attention. Thus, infants at this age begin to flexibly and reliably look where adults are looking (gaze following), use adults as social reference points (social referencing), and act on objects in the way adults are acting on them (imitative learning)—in short, to 'tune in' to the attention and behavior of adults toward outside entities. At this same age, infants also begin to use communicative gestures such as the pointing gesture to direct adult attention and behavior to outside entities in which they are interested—in short, to get the adult to 'tune in' to them (Tomasello, 1995).

In a large-scale longitudinal study, Carpenter, Nagell, and Tomasello (1998a) found that this whole panoply of joint attentional skills (measured by 9 different tasks) emerged in all children studied in close developmental synchrony, in correlated fashion, and with a highly consistent ordering pattern across children reflecting the different levels of specificity in joint attention required. One hypothesis is that these many different skills of joint attention emerge in developmental synchrony because they are all manifestations of a single underlying social-cognitive skill, namely, the understanding of persons as intentional agents who have a perspective on the world that can be followed into, directed, and shared (Tomasello, 1999). Support for this hypothesis comes from studies of how one-year-olds understand the behavior and perception of other persons.

In terms of the understanding of behavior, human infants' unique skill is their understanding of intentional action since this involves understanding something of the mental dimension of behavior—the differentiation of the actor's actions, her means, from her mental representation of the end state at which she is aiming, her goal. In preferential looking and habituation paradigms infants show some sensitivity to some of the properties of goal-directed action by the second half of the first year of life, although it is doubtful that this sensitivity indicates that the babies differentiate means and goals (Gergeley, Nadaszy, Csibra and Biro, 1995; Woodward, 1998; Baldwin and Baird, 2001). More clearly, when one year old infants attempt to imitate the goal-directed actions of others in overt behavior, they re-enact the action and simultaneously look in anticipation to the goal-object (Carpenter *et al.*, 1998a), and they even can evaluate why an adult chose the behavioral means she did rather than another (e.g., she chose an unusual means because the normal means were blocked; Gergeley, Bekkering and Király, 2002). Also, when 18-month-olds see an adult trying to do something they reproduce what she was

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

trying to do not what she actually did, implying an ability to infer the intentions underlying an action even if they were not actually consummated in perceptible behavior (Meltzoff, 1995; Bellagamba and Tomasello, 1999). Further, 16-month-old infants preferentially imitate intentional over accidental actions (Carpenter, Akhtar, and Tomasello, 1998b), demonstrating an ability to interpret basically "the same" behavior in different ways (as a goal-directed action or as an accident). And finally, 24-month-old children even understand prior intentions in the sense that they interpret the exact same behavior differently depending on how they understand the adult's intention as indicated in the moments immediately preceding the target behavior; for example, if an adult pulls at a box before engaging in some actions leading ultimately to opening it, young children construe the entire sequence as 'trying to open the box' in a way that they do not if they do not see the initial pulling (Carpenter, Call and Tomasello, in press, a). One to two year old children understand the basics of intentional action.

In terms of the understanding of perception, the key skill of human infants is an understanding not of perception in general—which may be shared with other primates (Tomasello, Call and Hare, 1998; Tomasello, Hare and Agnetta, 1999)—but of attention more specifically. Understanding another person's attention also bears the mark of the mental in that it involves knowing that persons have intentional control over their perception and that in particular cases they can choose to focus on one aspect of a situation rather than others that are also currently perceptible. In one of the only studies investigating infants' understanding of attention, Tomasello and Haberl (2002) had infants at 12 and 18 months of age play with two adults and two new toys. Then one of the adults left the room while the child and the other adult played with a third new toy. The first adult then returned, looked globally at all three toys aligned on a tray and exclaimed excitedly 'Wow! Cool! Look at that one! Can you give it to me?'. To retrieve the object the adult wanted, children had to know that people attend to and get excited about new things, and also to identify which one was new for the adult, even though it was not new for them. Even 12 month olds were successful in this task (which also had a control condition), demonstrating a nascent understanding that within their perceptual fields persons may choose to focus their attention on some things to the exclusion of others. One to two year old children also understand some of the basics of attention.

By virtue of their understanding of the intentions and attention of other persons, one to two year old children are able to engage in joint attentional activities that illustrate the first two of our three key characteristics. First, they particpate in joint attentional activities that are 'shared' in that they require that the child make some sort of self-other equivalence (Baressi and Moore, 1996; Tomasello, 1999; Hobson, 2002). For example, to attempt to draw someone's attention to something I am already focused on, so that we may share interest and attention to it, I must understand that the two persons involved may be focused on the same or different things. Similarly, to imitate someone's intentional action, I must understand that there are two persons involved—someone else and myself—who can perform the same goal-directed action. Second, one to two year old children also participate in

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

some joint attentional activities that require them to appreciate the notion of an attentional or mental perspective or description. This is most clearly apparent as they begin to make active choices about how to construe things linguistically—this is a dog, an animal, a pet, a pest, or even 'it'—for purposes of interpersonal communication. And it is not just that these perspectives are elicited from children differently on different occasions; children sometimes even use one and then immediately self-correct to another in the same breath ('the man. . . . the policeman'; Clark, 1997). It is thus clear in such cases that the child is choosing from among two or more descriptions that she knows are simultaneously available both to herself and to her interlocutor—and that they both know the descriptions not chosen (which enables many Gricean inferences).

The understanding of persons as intentional agents at one to two years of age—in ways that show sharedness and perspective—thus inaugurates the development of uniquely human skills of social cognition. This understanding involves appreciation of the 'original normativity' constitutive of actions in the sense that an intentional agent's action—either self or other—may be judged as successful or unsuccessful. One-year-olds have thus entered, at least in a nascent and implicit way, the space of reasons involving normative judgements. But of even more importance in the current context, children of this age also come to appreciate that shared intentionality and collective practices create 'derived normativity'—a more deeply social sense of normativity pertaining to the use of symbols, artifacts, and other culturally constituted entities. These entities are invested with normativity through the actions of intentional agents and their attitudes: this is the way 'we' use this symbol or tool; this is the way it 'should' be used; this is its 'function' for us, its users. Appreciating derived normativity is thus our third key characteristic making one year old cognition 'the real thing', and it is most readily apparent in the use of linguistic symbols and material artifacts such as tools and toys.

1.2. Learning and Using Linguistic Symbols

Human infants begin to show species unique communicative behavior during the nine-month revolution, before they have learned any language. Specifically, human infants begin to actively direct the attention of other persons to outside objects and events, for example, by pointing to them or holding them up and showing them to others, solely for the purpose of sharing attention. These behaviors—from the point of view of both production and comprehension—indicate that infants not only understand intentions but also communicative intentions. No other species on the planet attempts to direct the attention of others by pointing or showing outside objects in human-like ways, and so arguably no other species understands these kinds of communicative intentions.[3]

[3] Some apes raised by humans learn to point for things they want, but only for humans and not just to share attention (Tomasello and Camaioni, 1997).

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

But pointing and showing are only very generic attention directors, not adapted for particular referential situations. In contrast, linguistic symbols are social conventions that have evolved historically for directing attention in specific ways, that is, for inducing others to construe, or take a perspective, on some experiential situation. For example, in different communicative situations one and the same object may be construed as a car, a vehicle, or an SUV; one and the same event may be construed as running, moving, fleeing, or surviving; one and the same place may be construed as the coast, the shore, the beach, or the sand—all depending on which aspects of the shared experience the speaker wishes to draw the listener's attention to. As the child masters the linguistic symbols of her culture she thereby acquires the ability to adopt multiple perspectives simultaneously on one and the same perceptual situation, typically choosing to linguistically express just one of these in any given situation but sometimes more (Clark, 1997; Tomasello, 1999).

The other, more basic, thing about linguistic symbols is that they are intersubjective (bi-directional in the sense of Saussure, 1916)—meaning that they are comprehended and understood in the context of self-other equivalence. Assuming a child who can understand the adult's communicative intentions—that is, understand that the adult is making that sound with the intention that I share attention to X—a symbol is created when the child then acquires the appropriate use of the symbol herself. To do this she must understand that when she wishes to do as the adult is doing—when she wishes to get the adult to share attention to X—she may use this same sound. This form of cultural (imitative) learning thus differs from those in which the child imitates an adult action on an object directly in that there is a role reversal involved: the child uses the new symbol to direct another person's attention precisely as they have used it to direct her attention (the role reversal comes out especially clearly in deictic terms such a *I* and *you*, *here* and *there*). The child's use of the same sound as the adult, for the same purpose, thus creates a communicative convention, or symbol, that the child produces and at the same time appreciates that the recipient comprehends and might potentially produce (Tomasello, 1998). We may think of this bi-directionality or intersubjectivity of linguistic symbols as simply the quality of being socially 'shared'.

But what about normativity? What evidence do we have that young children view linguistic symbols reflectively and normatively? The major evidence is children's tendency in the second year of life to play with words and how they are used, in a manner very similar to symbolic play with objects (to be discussed in more detail below). Thus, with a child approaching her second birthday one can systematically misname objects in a playful way, for example, calling an elephant a giraffe, and they will sometimes join into this game—both laughing at the adult play with words and contributing themselves (Clark, 1978; Horgan, 1981; Johnson and Mervis, 1997). As we will argue in more detail below, this kind of play with the conventional use of things—in a way that clearly indicates the child's understanding of the convention and it's breaking—illustrates that children participate in the

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

use of symbols with some kind of reflective understanding of their conventional/normative dimension, how one ought to use them under normal circumstances. But this reflectivity is much more readily apparent in the use of material artifacts because they can have, much more easily than relatively evanescent linguistic symbols, multiple functions.

1.3. Learning and Using Artifacts

From 3 or 4 months of age, human infants are interested in objects and so begin grasping, banging, and sucking them. Many of the objects infants interact with are artifacts pre-fashioned in some way by adults, but at the beginning they are not recognized as such—they too are grasped, banged, and sucked. But as infants approach their first birthday, as part of the 9-month revolution, they begin to appreciate the intentional dimension to artifacts, that is, their specific functions. Although on occasion the child may discover the function of an artifact via its own individual explorations, in general the intentional dimension of an artifact comes into being as the child observes other persons using it. But this may happen in one of two ways, depending on the nature of the social learning processes involved—and this makes a difference to the child's understanding.

The two types of social learning are emulation learning and imitative learning more strictly defined (Tomasello, 1990, 1996). Emulation learning is a form of social learning that does not rely on the observation and reproduction of the goal-directed behavior of other persons. In emulation learning an observer watches someone manipulate an object and learns something new about the object as a result, which it may then use to devise its own behavioral strategy. For example, one primate might crack open a nut that an observing conspecific did not previously know was a food item that could be opened. The observer thus learns 'that object is a food item that can be opened' and so proceeds to try to figure out a way to crack open the nut for itself, with no attention to the strategies used by the original nut cracker. Emulation learning is the major way in which nonhuman primates learn about their environments in social situations, and indeed it also plays a major role in human infants' initial explorations of many artifacts (von Hofsten and Siddiqui, 1993). In emulation learning children learn what objects do.

The second type of social learning is imitative learning in which an observer attempts to copy the goal-directed behavioral strategies of others—a type of social learning that may be uniquely human (although there is much controversy on this point; see Tomasello 1996; Call and Carpenter, 2002). This does not mean that the observer blindly mimics the sensory-motor actions of others—the way that a parrot mimics human speech, for example—but that the observer attempts to reproduce the intentional actions of the other, including the goal toward which they are aimed—as illustrated most clearly in the studies (described above) of Carpenter *et al.* (1998a) in which 12-month-olds anticipate goals, Meltzoff (1995) in which 18-month-olds reproduce what an adult is trying to do, and Carpenter *et al.*

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

(1998b) in which 16-month-olds reproduce intentional but not accidental actions. This kind of cultural learning requires the understanding of other persons and oneself as intentional agents, which brings the goal-directed actions of other people and one's own actions under the same description ('I do what you did')—thus fulfilling the Generality Constraint (Evans, 1982) on simple concepts of persons and actions.

Tomasello *et al.* (1993) attempted to capture the essential difference between these two major types of social learning by saying that in emulation the observer learns *from* the demonstrator, whereas in imitation (one form of 'cultural learning') the observer learns *through* the demonstrator—understanding the intentional structure of the demonstrator's behavior and then trying to do what she is doing. By engaging in this interpretive process while observing an adult using a symbol or artifact, the child learns what 'we', the users of the symbol or artifact, do with it—what it is 'for', its physical function: 'this object can be used to do X in context C' (Searle, 1995). This gives the artifact a kind of derived normativity—tools can be said to be working well or badly, or can be used appropriately or inappropriately. Infants thus come to pick up the physical functions of artifacts assigned to them by shared intentionality via cultural (imitative) learning. Interestingly, recent research has shown that children's initial understanding of object functions may be tied to what they see being done with them at the moment by specific people, and that their understanding of what objects are 'for' in the culture more generally—that is, their ability to take the so-called design stance—develops gradually over the preschool years (Bloom and Markson, 1998; Matan and Carey, 2001; German and Johnson, 2002).

Infants' introduction into the collective practice of assigning functions to objects becomes even clearer in a phenomenon of late infancy known as pretend (or symbolic or imaginative) play. Sometimes infants and young children do not use artifacts in instrumental, physically functional ways, but instead—in concert with an adult—play with the object's function in creative ways. Thus, a 2-year-old might pick up a pencil and pretend it is a toothbrush. But as Hobson (1993) has pointed out, the child is doing more than simply manipulating the pencil in an unusual way. In pretend play the infant also looks to an adult with a playful expression: she knows that this is not the proper function of this object and that her unconventional use is something that may be considered 'funny'. An act such as this very clearly involves the child in a perspective shift, anointing the object with a new, temporary description. This process can be understood as shared assignment not of a physical function to the object—because the pencil clearly does not serve as a real toothbrush—but of a status function (Searle, 1995): 'This object counts as a toothbrush in our pretense context'. The shared intentionality involved in this creation of status functions is of a stronger kind than in the assignment of physical functions. In the case of physical functions one makes use of intrinsic causal properties of objects and uses them for specific practical purposes—which makes them 'tools'. In the case of status functions in pretense, one treats objects collectively as if they were something else, virtually irrespective

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

of their causal properties and without concrete instrumental purposes—which makes them 'toys'. The pencil counts as a pretend toothbrush only because it is collectively treated as such in the pretense episode. Accordingly, the role of cltural (imitative) learning should be even stronger in learning to use 'toys' than in learning to use 'tools'.

Recent research has provided evidence for this interpretation of pretend play. The first point is that although it is widely assumed that children's early pretend play is constituted by acts of individual creativity, this is in actuality not the case (at least there is no evidence for it). In an experimental study, Striano, Tomasello and Rochat (2001) provided 18 to 30 month old children with various opportunities for using various kinds of objects and toys symbolically. But children below 24 months of age almost never produced a creative pretense act with an object that they had not seen another person use symbolically first. (This finding is also consistent with some informal reports that children living in cultures in which there are few toys, and in which adults do little to model or encourage pretend play, engage in very little pretend play with objects themselves; J. Linaze, P. Brown, personal communications.) When children over two years of age imitated a pretense act in this study, they tended to look more at the experimenter than when they imitated an instrumental act (and in some cases to smile more as well)—perhaps evidencing that they were beginning at this age to understand something of the shared intentionality and different descriptions that went into the creation of the pretense reality.

Extending this line of investigation, Rakoczy, Tomasello and Striano (2002) attempted to simulate children's initial encounters with tools and toys by providing them with a set of totally novel objects. Some of the objects were demonstrated to have instrumental functions, whereas others were demonstrated to have pretense functions. Over three encounters, children began doing with these objects what adults did with them. Overall, it was found that the instrumental demonstrations were easier to imitatively learn, and children generalized these more readily to other objects. Children imitatively learned the pretense demonstrations also, but they did not generalize these creatively to other objects until 24 months of age. The argument was thus that children imitatively learn the functions of objects—what we do with them—in very similar ways for tools (artifacts used instrumentally) and toys (artifacts used in pretense). But the adult intentions behind these two kinds of acts and the functions they create are different, and at around two years of age children begin to perceive this (again as indicated by looking and smiling to the adult): tools are used to causally effect concrete sensory-motor ends, whereas toys are used to engage in a special kind of shared intentionality in which we together create a new function. We may thus say that learning to use tools is *socially mediated*, since children learn the intentional affordances of these artifacts through adults (but could potentially learn the causal properties of tools on their own), whereas learning to use toys for pretense purposes is *socially constituted*, since adults and children create the functions on the spot. The shared intentionality in pretense thus constitutes a special kind of derived normativity: this pencil is temporarily

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

a toothbrush (its status function), and this joint declaration commits us to interacting with it in certain ways (see Currie, 1998).[4]

1.4. Shared Intentionality in the Second Year of Life

Our argument is thus that already at one to two years of age young children have begun to engage in uniquely human forms of social cognition. Virtually all of their joint attentional activities require them to make some kind of self-other equivalence, leading to activities that are 'shared', and in some of these they take the perspective of other persons, sometimes showing the ability to knowingly provide different descriptions of the same phenomenon. In addition, as young children begin to interact with public artifacts, they demonstrate a kind of reflective understanding of the social-normative dimension of these special cultural entities.

In the case of language, children learn to use linguistic symbols in shared practices, exploiting their bi-directional nature, and to apply them to objects in context-sensitive ways, thereby establishing different perspectives (descriptions) on one and the same entity. Moreover, children can play with the normal, conventional use of symbolic artifacts such as words, and be amused by that, in much the same way they play with the normative uses of material artifacts. In the case of children's pretend play with material artifacts, they initially learn to act on objects symbolically by imitatively learning adult acts of pretend play, employing a self-other equivalence, and of course the defining quality of pretend play is the provision of not-normal, temporary descriptions of things. But pretend play also involves a kind of shared intentionality in which we (child and adult) conspire to create a new function for an object that we both know together and reflectively is not its 'normal' function. By participating in activities with symbolic and material artifacts displaying sharedness, perspectivity, and derived normativity, children begin to enter in earnest into the collectivity that is human cognition.

But two year old children's understanding of what they are doing is not the same as that of 4 and 5 year old children's. Two-year-olds participate in collective practices and reflectively understand, in some sense, the intentional perspectives embodied in shared actions and the derived normativity they confer on objects ('this is what we do with this object'). They can thus be said to have internalized intentional social perspectives which they use as reference points in dealing with objects. But across development the nature of these social perspectives changes. We propose that from 1 to 4 years of age children go from participating in shared intentionality involving the internalized perspectives of other specific individuals in specific action contexts, such as a parent or sibling in a pretend game of 'brushing teeth', to those characterized by collective intentionality, in which they appreciate

[4] We should note that this same analysis applies to representational toys such as toy dolls and cars. At first children do not comprehend their iconic status at all, and so imitatively learn to manipulate them like adults. Later they use them in pretense, with the iconic dimension perhaps aiding the process (although we know very little about this empirically).

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

and utilize in all situations the more generalized and abstract set of perspectives and norms—often instantiated as 'beliefs'—characteristic of the culture as a whole.[5]

2. Understanding Mental Agents

One and two year old children thus know a lot about other people. They know important things about what others see, do, intend, and attend to. They know some things about the intentional structure of language and the way it can provide different perspectives and descriptions of things, and also about the intentional structure of material artifacts and the way one can play with this intentional structure. They know that the use of symbolic and material artifacts is conventional and normative, such that one can share mirth at violations of use.

With all of this in place before the second birthday, a reasonable question is: what is missing? Why do young children not begin to operate with a full-fledged belief-desire psychology for another two years or more? What is so difficult about understanding that people's behavior is directed by what they believe is the case and not what really is the case? And despite some criticisms of traditional tasks for assessing children's understanding of false beliefs, in a recent meta-analysis Wellman, Cross and Watson (2001) found that the many attempts to make the tasks more child-friendly have resulted in only minor improvements in children's performance. The age is still 4 to 5 years. Call and Tomasello (1999) even developed a nonverbal version of the task (which correlated well with the verbal version), and found children passing at around the same age. In fairly drastic modifications of the task, some researchers have found that children can deal with other people's beliefs at around their third birthdays (Clements and Perner, 1994; Carpenter *et al.*, in press, b), but it is not clear that these tasks require the same level of understanding beliefs.

There are of course many possible answers to the question of why children find it so difficult to understand beliefs, and the truth is there simply is not enough empirical research for anyone to feel confident about an answer. But our proposal for the moment is that to understand beliefs young children must learn to differentiate—in a way that one- and two-year-olds cannot yet do—between the

[5] Our distinction between shared intentionality and collective intentionality is similar to Searle's (1995) distinction between collective intentionality broadly understood (yielding social facts) and collective intentionality proper involving constitutive rules and the creation of institutional facts. However, the two distinctions do not match perfectly: we contend that in shared intentionality 1-year-old children may actually create a socially defined product (e.g., in pretend play with others), and these share important features with institutional facts. What changes after four years of age is that children become able to participate in and understand facts created not just by themselves and a partner in a momentary interaction, but rather those created by the culture at large through a system of beliefs and practices. [We should also note that we disagree with Searle's claim that hyenas hunting together show shared intentionality; we contend that shared intentionality is a uniquely human phenomenon.]

UNCORRECTED PROOF

© *Blackwell Publishing Ltd. 2003*

mental perspective of an individual and 'reality'. And reality is not just the child's individual perspective of the moment, which may conflict with another person's, nor an intersubjectively shared perspective with other persons, but rather it is objective in the sense that no one perspective is privileged (the view from nowhere). The notions of objective reality, subjective beliefs, and intersubjective perspectives thus form a logical net that can only fully be grasped as a whole. Comprehending this net as a whole takes children, apparently, several years to accomplish.

2.1. The Role of Language

Following a growing number of researchers, we believe that a critical role in children's construction of a belief-desire psychology—understanding persons as mental agents—is played by processes of linguistic communication. Thus, a number of studies have found significant correlations between children's linguistic skills and their skills in false belief tasks, even when the language measures are taken one or two years before these tasks. Studies of this kind are reported by Dunn, Brown and Beardsall (1991), Astington and Jenkins (1995, 1999), DeVilliers and DeVilliers (2000), and Farrar and Maag (2002), with some correlations in the .60 to .70 range. Relatedly, Peterson and Siegel (2000) have found that deaf children who grow up with deaf parents fluent in sign language, and who therefore have fairly normal linguistic experience, are significantly better at false belief tasks than other deaf children who grow up with hearing parents whose relatively poor sign language skills means that their children have impoverished linguistic experience.

But in none of this research is it possible to tell with any degree of specificity which aspects of linguistic experience are most important or crucial. For example, is the crucial factor adult use of linguistic symbols to indicate mental states such as *think*, *know*, and *believe* (Bartsch and Wellman, 1995; Astington and Jenkins, 1995, 1999)? Or is it the syntax of the way adults talk about beliefs and related mental states (i.e., in sentential complement constructions) that provides children with a necessary, or at least facilitative, representational format for dealing with mental concepts cognitively (DeVilliers and DeVilliers, 2000)? Or is the key the process of discourse in general in which the child comes to appreciate that other people know things she does not know and have different perspectives on things than herself and other people (Harris, 1996, 1999; Tomasello, 1999)?

In an attempt to identify the effective factors more specifically, Lohmann and Tomasello (in press) trained three year old children in one of four different training conditions involving adult-child interactions with deceptive objects (e.g., children see an object that looks like an apple but is 'really' a candle) and various kinds of accompanying language (including one condition with no substantive language). The outcome measure, taken after 3 training sessions, was several different types of false belief tasks. There were three main findings. First, and perhaps most importantly, simply experiencing deceptive objects was not enough to facilitate children's false belief understanding; some language from other persons structuring that

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

experience seemed to be necessary. Second, two types of linguistic experience were found to be most facilitative: (i) an adult pointed out in discourse with different words the different possible perspectives on the objects (e.g., *apple*, *candle*), and (ii) an adult used many utterances with sentential complement constructions, including mental state predicates (propositional attitudes) as matrix verbs—even without any experience with deceptive objects. Third, these two effects—of perspective-shifting discourse and sentential complement syntax—seemed to be relatively independent of one another, as the strongest facilitator of children's false belief understanding in this study was a training condition incorporating both of these factors.

If indeed it is the case that these two factors—perspective-shifting discourse and sentential complement constructions—each plays an important and independent role, we should look at each a bit more carefully.[6]

2.2. Perspective-Shifting Discourse

Young children acquire and use their language in discourse from the beginning of language development in the middle of the second year of life. But sometime after their second birthdays many children begin to command linguistic skills advanced enough to enable them to engage in more sophisticated discourse interactions with a real give-and-take of perspectives, that is, those involving not just the different perspectives implicit in the use of linguistic symbols and constructions, but the explicit perspectives that interlocutors linguistically express toward one another in propositions—sometimes concerning one another's previously expressed propositions. As children engage in such discourse, they are constantly simulating the perspective of the other person and relating that to their own perspective (Harris, 1996; Tomasello, 1999). There are several forms of discourse that seem especially important in children's coming to understand the notion of a mental agent.

One especially important form of discourse in this context is disagreements and misunderstandings. Thus, children have discourse interactions with some regularity in which one person expresses the view that X is the case, and the other disputes this and claims that Y is the case. Or, similarly, interactants may have a clear difference of knowledge or beliefs as when the child makes a presupposition that the other does not hold in kind (e.g., the presupposition of shared knowledge in using *He* or *It*), or the same thing may happen in reverse as other persons make unwarranted presuppositions about shared knowledge and beliefs they have with the child. Also important may be (a) misinterpretations, in which the adult interprets the child's utterance in a way that she did not intend, and (b) clarification requests in which the child says something that the adult does not understand and so the adult asks for clarification. These situations lead the child to try to discern why the adult does not comprehend the utterance—perhaps she did not hear it, perhaps she is not familiar with this specific linguistic formulation, and so forth.

[6] See Lohmann (in preparation) for a more in depth discussion.

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

In all, it would seem that these kinds of disagreements, misunderstandings, and repairs are an extremely rich source of information about how one's own understanding of a linguistically expressed perspective on a situation may differ from that of others.

But perhaps of most crucial importance—especially for the normative component of the understanding of mental agents and their beliefs—is reflective discourse in which the adult and child comment on the ideas contained in the discourse turn of the other. For example, a child may comment on the way she is doing something, and her father may then give her an instruction for a better way to do it. In the hypothesis of Vygotsky (1978), this instruction is then internalized and the child uses it to self-regulate her subsequent actions in that context (see Kruger and Tomasello, 1986; Kruger, 1990). What is internalized is something normative: the adult's evaluation of the child's expressed thought, which by including an evaluative component encompasses that thought. Perhaps because this is all done in the common cultural symbols of a natural language, the evaluation may potentially be represented as the perspective of the culture as embodied in the adult's voice (more on this below). Perhaps not coincidentally, children show relatively clear evidence of internalizing adult regulating speech, rules, and instructions as they are reaching age 4 to 5 and beginning to solve false belief tasks.

The specific hypothesis is thus that the transition to an understanding of mental agents is a gradual process that derives at least partly from the child's use of intentional understanding in discourse in which there is a continuous need to take into account other persons' perspectives on things as expressed in propositions, which often differ from the child's own perspective. Of special importance may be reflective discourse, which at least potentially can convey normative perspectives emanating from the culture at large. The internalization of these dialogic interactions leads to a reflective stance, incorporating both cultural norms and, as differentiated from these, individual beliefs. Importantly, beliefs differ from simple perspectives in that beliefs involve a commitment to truth that can guide action. Perhaps understanding commitment in believing emanates from appreciating commitment in asserting, which becomes especially clear in reflective discourse, with its need to justify and to stand up to the dialogical challenge of other people's evaluations of one's own assertions. And because the interlocutors are all using the same culturally conventional language, these evaluations may come to constitute the collective background reality that forms the context for the statement of individual beliefs.

2.3. Propositional Attitude Constructions

Children's increasing mastery of their native language during the period from age 2 to 4 also includes mastery of a special class of syntactic constructions known as sentential complements. These constructions prototypically have some kind of psychological verb expressing a propositional attitude as the main verb (e.g., *say*, *know*, *think*, *believe*) and then a full proposition indicating what someone says,

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

knows, thinks, or believes. DeVilliers and DeVilliers (2000) have emphasized that these constructions provide a handy, perhaps even necessary, representational format for children to cognitively represent such things as beliefs (in all of their referential opacity). An important role may also be played by the semantics of the psychological verb, and indeed the DeVilliers think that epistemic verbs, such as *know* and *believe*, are most important as they indicate most directly the requisite propositional attitude (with referential opacity).

Sentential complement constructions thus symbolically indicate propositional attitudes, and being able to conceptualize these as distinct from the propositions they encapsulate requires a certain type of pragmatic understanding of the way linguistic communication works. Thus, in mature linguistic communication speakers monitor two main things. First, they monitor what they want to say, the basic who-did-what-to-whom they want to report (the proposition). But second, they also monitor the knowledge and expectations of the listener and so formulate their proposition in ways appropriate to the immediate speech situation, for example, using pronouns for shared information, using a relative clause to disambiguate a referent, using a passive to efface the agent of an action, or using a modal or psychological verb to indicate the speaker's attitude. Initially for young children these two tasks are not differentiated; they simply comprehend and use constructions they know fairly indiscriminately, as prompted by various discourse situations. But with greater experience children begin to see a difference between the propositions expressed in the conventional symbols of language and the pragmatic choices and adjustments made by individual persons on individual occasions of language use. The propositional attitudes actually encoded in language for use on specific occasions—for example 'I think . . .'—give children a handy way to get some reflective purchase on this differentiation.

Importantly, mastery of propositional attitude constructions (sentential complements) is a gradual process. Thus, Diessel and Tomasello (2001) followed in longitudinal detail five children's mastery of these constructions. What they found was that all children began with a small set of formulaic expressions of propositional attitudes, such things as *I think*, *I bet*, *You know*, *I hope*, and so forth (see also Bartsch and Wellman, 1995). Some of these are so formulaic they could even be replaced easily by an adverb like *maybe* (I think X = Maybe X). The children almost never at this stage used propositional attitude verbs in the past tense, with a negative or other modal, or with a third person subject; a given child's use of a particular verb was practically invariant. But then gradually from about age 2 1/2 to age 3 1/2 they began to use a much more varied set of ways to indicate propositional attitudes of different types containing different verbs, third person subjects, modal operators such as negatives, different tenses, and so forth.

One way to explain the process is this. In learning to comprehend and use an ever wider array of propositional attitude clauses, the child is linguistically bootstrapped from the *expression* of propositional attitudes to the understanding and *ascription* of propositional attitudes to herself and others. In the case of beliefs, the child first learns that 'whenever you are in a position to assert that p, you are ipso

© Blackwell Publishing Ltd. 2003

UNCORRECTED PROOF

facto in a position to assert "I believe that p"' (Evans, 1982, p. 225f), but without having any conceptual understanding of beliefs as such. Likewise, the child learns, first without full understanding, that he can say 'X believes p' and that X herself can say 'I believe p' whenever she asserts p, and so forth. More complicated procedures include learning to say 'You believe p, but not p' instead of simply 'No! Not p' whenever the child disagrees with the interlocutor. Of course, much more beside these purely formal procedures is needed for the child to acquire a concept of belief: she also has to learn to use such things as 'I believe' and 'X believes' in reason-giving discourse, which provides her with the raw material for constructing the complex interrelations among the concepts of belief, reason, and truth. And so, using propositional attitude clauses provides a formal bootstrapping device—not sufficient but probably necessary—for understanding adult-like concepts of rationality. The child acquires full command of the adult-like concepts—including the ability to coordinate first person usage, without application of criteria, and third person usage, on the basis of behavioral and other linguistic criteria (and so to fulfill the Generality Constraint on psychological predicates; Evans, 1982; Strawson, 1959)—only gradually as she uses them in an ever wider variey of discourse contexts, especially those involving the giving and demanding of reasons.[7]

2.4. Collective Intentionality in the Third and Fourth Years of Life

Two year old children, we have argued, engage in joint attentional activities and use symbols and artifacts in ways that evidence their understanding of self-other equivalence, perspectivity, and normativity. But when interacting with other persons what they are capable of dealing with is only perspectives expressed implicitly in language (*dog* vs. *animal, chase* vs. *flee*), not with explicitly stated beliefs. Similarly, the norms two year old children are capable of dealing with are those implicit in the use of symbols, artifacts, and cultural conventions as they interact with other specific individuals. For the 2-year-old, these norms are about what you and I should do with this artifact or symbol right now—how you and I use it—and so the voice from which the norm comes is particular individuals on particular occasions.

But over the course of two years of relatively continuous dialogical interaction with other persons, young children are confronted with all kinds of perspective-shifting discourse, including reflective discourse, and with propositional attitude constructions.

[7] This account is broadly consistent with elements of Gordon's (1995, 1996) simulation account in which children first learn to follow so-called ascent routines: Evans' Procedure (1982) for beliefs and analogous ones for other mental states (e.g., in the case of desire to say 'I want p' instead of 'P would be nice'), without having the relevant concepts (e.g., of belief or desire) in full-fledged form. Based on this practice children then come to use propositional attitude constructions in a truly ascriptive way by embedding ascent routines in simulations of other persons, thereby learning to affix to 'p' not only 'I believe', but also 'You believe', 'She does not believe', and so forth. The outcome is that explicit concepts of mental states emerge gradually out of the child's expressive linguistic practices and non-conceptual procedures.

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

Reflective discourse in particular is important because it represents cases in which an interlocutor expresses an attitude toward the child's perspective (either linguistically expressed or not), and so it is a kind of inter-personal reflectivity. Propositional attitude constructions also represent a kind of reflectivity, in this case mostly intra-personal, as the child registers both a proposition and its encapsulation in the wider perspective or attitude of the speaker—all in one representational package. And indeed, one can imagine that propositional attitude constructions emerged historically as rhetorical moves in discourse, especially reflective discourse, in which interactants must constantly mark their own attitude and those of others towards the same proposition(s). Because all of this is done in the conventional practices of a natural language, and so is effected in similar ways by many different individuals in the child's experience, over time the child may come to understand and internalize a kind of culturally general set of perspectives, assumptions, and norms. These provide the objective background reality against which beliefs can be explicitly ascribed to self and other individuals in ways that may or may not match this reality. Natural language with its recursive structure thus enables the child first to commit herself to truth in asserting and then to reflectively refer to these commitments in discourse and belief.

We may thus say that 2-year-olds participate in shared intentionality with specific other persons, whereas 5-year-olds participate in collective intentionality with individuals representing a broader set of cultural perspectives and norms. In the terms of Mead (1934), the child is going from guiding its actions via an internalized 'significant other' to guiding its actions via an internalized 'generalized other'. Importantly, this difference enables a new understanding of human mental activity in terms of not only individual beliefs but also of collectively intentional beliefs—which have the world-making power to create cultural-institutional realities. Thus, 2-year-olds' understanding of intentions simply does not enable them to grasp the workings of cultural institutions such as money, marriage, and government—whose reality derives from collective practices and beliefs in their existence—whereas 4- and 5-year-olds, with their newly acquired concepts of belief and reality, are in a position to begin learning about these collective entities. Indeed in virtually all cultures in which there is formal education, where children learn about such things as cultural institutions and their workings, 5 to 6 years of age is the canonical starting point (Kruger and Tomasello, 1996).

The process by which 2 to 5 year old children acquire the logically interwoven notions of objective reality, subjective beliefs, and intersubjective perspectives can thus be described as one of representational redescription (Karmiloff-Smith, 1992). The psychological abilities that enable 2-year-olds to engage with others in linguistic and other conventional practices—with shared, conventional (normative) symbols embodying intentional perspectives—become first expressed in language and then redescribed in that very same language. Engaging in linguistic communication and discourse with other persons—in which some of the discourse is about the content of previous discourse—thus enables a developmental progression from the expression of one's own perspective and the practical coordination of

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

multiple perspectives to the explicit ascription of potentially different perspectives, indeed beliefs, to oneself and other persons.

3. And What About the Apes?

Everyone knows and agrees that human beings are the only species to engage in collective intentionality of the type needed to create such things as money, marriage, and government. But this is like saying that only human beings build skyscrapers, when the fact is that only human beings, among primates, build any shelters at all. If we want to get to the essence of human cognitive uniqueness, therefore, the focal point should be shared intentionality, since that is what underlies more basic, but still unique, human skills such as language, cultural learning, and pretense.

Until recently, differentiating human and other ape social cognition was relatively easy. Although there were some reported anecdotes of apes doing theory of mind like things, there were no convincing experimental demonstrations that they could understand any psychological states of other beings at all, and there were a number of negative findings (see Tomasello and Call, 1997, Povinelli *et al.*, 2000, for reviews). However, two more recent lines of research suggest that modifications of that conclusion are needed—and the nature of those required modifications is instructive.

First, it turns out that apes may understand something about intentions. Call, Hare, Carpenter and Tomasello (2002) presented chimpanzees with a human who had food in his hands and then behaved in different ways indicating that he was either unwilling or unable to give them the food. There were three conditions in which the experimenter was unwilling in different ways (e.g., just staring at the ape, eating the food, teasing the ape with the food). These conditions were each paired with two unable conditions (e.g., trying to get the food out of a jar, dropping it accidentally, etc.). In each group of matched conditions the topography of the experimenter's behavior (body movements and gaze direction) were kept as similar as possible. The main finding was that chimpanzees were more impatient—banged on the cage more, left the area sooner—when the human was being intransigent (unwilling) than when the human was making a good faith effort (unable), even though in neither case did they get the food. Importantly, the findings were strongest in those conditions in which the experimenter specifically acted on the food—e.g., had an accident with it or used to tease the ape—as opposed to conditions in which there was little action—e.g., the experimenter just sat there or was distracted away from the food. This is important because it means that the cue the apes used for identifying intentional behavior was perceptible in it—Searle's intention in action (see also Call and Tomasello, 1998, for some similar findings).

Second, it turns out that apes may also understand something about perception. Hare, Call, Agnetta, and Tomasello (2000) placed a subordinate and a dominant chimpanzee into rooms on opposite sides of a third room. Each had a guillotine door leading into the third room which, when cracked at the bottom, allowed

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

them to observe two pieces of food at various locations within that room—and to see the other individual looking under her door. After the food had been placed, the doors for both individuals were opened and they were allowed to enter the third room. The basic problem for the subordinate in this situation is that the dominant will take all of the food that it can see. However, in some cases things were arranged so that the subordinate could see a piece of food that the dominant could not see, for example, when it was on the subordinate's side of a small barrier. The question was thus whether the subordinates knew that the dominant could not see a particular piece of food, and so it was safe for them to go for it. The basic finding was that the subordinates did indeed go for the food that only they could see much more often than they went for the food that both they and the dominant could see. Several control procedures and conditions (one using a transparent barrier, that the subordinate apparently understood did not block the dominant's visual access to the food) effectively ruled out the possibility that the subordinate was simply monitoring the behavior of the dominant and reacting to that behavior.

In a follow up, Hare, Call and Tomasello (2001) used two barriers and one piece of food. In some trials the dominant witnessed the food being hidden behind one of the barriers, and the subordinate witnessed his witnessing. In other trials the dominant was absent for the hiding process, and the subordinate witnessed her absence. The main finding was that subordinates preferentially retrieved the food that dominants had not seen hidden, which suggests that subordinates were sensitive to what dominants had or had not seen during the baiting process. In an additional experiment, the dominant who had witnessed the baiting was switched (or not, in a control condition) for another dominant who had not witnessed the baiting before the competition began. Subordinates went for the food more often when the dominant had been switched than when she was not switched, thus demonstrating their ability to keep track of precisely who had witnessed what.

So it seems that apes can understand some psychological states in others, concerning both behavior and perception. With regard to behavior, the chimpanzees in the Call *et al*. study seemed to know something about intention in action. They apparently could see such things as effort, trying, frustration, and satisfaction, as signs of what the other person was about to do next. With regard to perception, the chimpanzees in the Hare *et al*. studies seemed to know what others could and could not see, and even what they had and had not seen in the immediate past. But what seems to be missing still is the shared dimension of all this. Chimpanzees might understand something about simple intentions and therefore even original normativity. But they seem not to understand communicative or cooperative intentions, and so they do not attempt to direct the attention of conspecifics by pointing, showing, offering, or any other intentional communicative signal (Call and Tomasello, in press).[8] And although they can learn to use human artifacts, apes

[8] Chimpanzee cooperative hunting is of the same type as that of lions. It is a complex social process but with relatively simple individual decision making (Cheney and Seyfarth, 1990).

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

do not engage in pretend play or any other behavior suggesting that they perceive the human intentionality and derived normativity embodied in those artifacts. Chimpanzees may also understand that conspecifics perceive things in their environment, but in this case again there is no evidence for the more deeply social, shared dimension of the process as we observe it in human children. For example, there is no evidence that in the Hare *et al.* studies the subordinate knew that the dominant was having first-person subjective experiences like her own only from a different perspective from across the cage—which would suggest an understanding of self-other equivalence and perspective. And of course apes do not use in their natural environments any kinds of shared symbols for taking perspectives on things. In all, there is basically no evidence in any sphere of ape activity that they can deal effectively with anything that is socially shared (self-other equivalence), perspectival, or normative.

One hypothesis is thus that apes understand the 'directedness' of others' behavior and can use various behavioral signs of effort and the like to make specific predictions on specific occasions about where that will lead next. They also understand when others do and do not see things, and have a memory for this and some knowledge of what it predicts about others' behavior. We might thus propose, in the direction of a proposal by Gergeley (2001), that chimpanzees and other great apes—and perhaps other primates and animals—possess a social-cognitive schema enabling them to see a bit below the surface and perceive something of the intentional structure of behavior and how perception influences it. Then on top of this schema—but actually woven in at a fairly early ontogenetic point—humans identify with other persons in ways leading to an understanding of self-other equivalence, which goes beyond this schema in leading to an appreciation of different social perspectives on things and ultimately to various kinds of derived normativity.

4. Conclusion

Our argument is thus that from an evolutionary point of view the one year old cognitive transition is the key one, and this transition does indeed concern the child's understanding of things mental. One to two year old children participate in cultural activities using shared, perspectival symbols with a conventional/normative/reflective dimension—the features of human cognition that constitute its rationality and at the same time so clearly distinguish it from that of other animal species.

Importantly, these early skills almost certainly have a specific biological basis. Children of all cultures begin engaging in joint attentional activities at around the same age, and there are no known environmental variables that significantly speed up or retard the 9-month revolution. Also, it is well known that the biologically based problems of children with autism begin with these one year old joint attentional skills; they do not wait for a four year old theory of mind to show

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

themselves (Hobson, 1993; Sigman and Capps, 1997). And these one year old understandings are what most clearly differentiate human cognitive abilities from those of other apes by enabling young children to participate in various forms of shared intentionality with other individuals. Finally, these one year old social-cognitive skills are the ones that underlie young children's earliest cultural learning and linguistic communication, key species traits by any account.

All of this is not true, or at least not so true, of the four year old transition. Children are reasonably variable in the age at which they undergo this transition, both within and between cultures, and there are known environmental variables that significantly speed up or retard children's understanding of mental agents—such things as family constellation, linguistic experience, and so forth. And there are no known biological deficits associated exclusively with the four year old transition not involving the 1-year-old transition. Our hypothesis is thus that this four year old transition depends on understanding persons as intentional agents (the one year old transition) as well as several years of linguistic interaction involving perspective-shifting discourse, the use of propositional attitude constructions, and reflective discourse. In the process children representationally redescribe their understandings of persons using the culturally shared symbols of their language, and so begin down the road not just of shared intentionally with other individuals but of the collective intentionality that constitutes their culture.

Max Planck Institute for Evolutionary Anthropology
Leipzig

References

Astington, J., and Jenkins, J. 1995: Theory of mind development and social understanding. *Cognition and Emotion*, 9, 151–165.

Astington, J., and Jenkins, J. 1999: A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, 35, 1311–1320.

Baldwin, D.A., and Baird, J.A. 2001: Discerning intentions in dynamic human action. *Trends in Cognitive Sciences*, 5(4), 171–178.

Baron-Cohen, S. 1993: From attention-goal psychology to belief-desire psychology. In S. Baron-Cohen, H. Tager-Flusberg and D.J. Cohen (eds.), *Understanding other minds: Perspectives from autism* (pp. 59–82). Oxford: Oxford University Press.

Baron-Cohen, S. 1995: The Eye Direction Detector (EDD) and the Shared Attention Mechanism (SAM). In C. Moore and P.J. Dunham (eds.), *Joint attention. Its origin and role in development* (pp. 41–59). Hillsdale, NJ: Lawrence Erlbaum.

Barresi, J., and Moore, C. 1996: Intentional relations and social understanding. *Behavioral & Brain Sciences*, 19(1), 107–154.

Bartsch, K., and Wellman, H.M. 1995: *Children talk about the mind*: New York, NY, US: Oxford University Press.

© *Blackwell Publishing Ltd. 2003*

UNCORRECTED PROOF

Barwise, J., and Perry, J. 1983: *Situations and attitudes*. Cambridge, MA: MIT Press.

Bellagamba, F., and Tomasello, M. 1999: Re-enacting intended acts: Comparing 12- and18-month olds. *Infant Behavior & Development*, 22(2), 277–282.

Bloom, P., and Markson, L. 1998: Intentionality and analogy in children's naming of pictorial representations. *Psychological Science*, 9, 200–204.

Call, J., and Carpenter, M. 2002: Three sources of information in social learning. In K. Dautenhahn and C. Nehaviv (eds.), *Imitation in animals and artifacts*. MIT Press.

Call, J., Hare, B., Carpenter, M., and Tomasello, M. (submitted) Unwilling or Unable? Chimpanzees' understanding of human intentions.

Call, J., and Tomasello, M. 1998: Distinguishing intentional from accidental actions in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*), and human children (*Homo sapiens*). *Journal of Comparative Psychology*, 112, 192–206.

Call, J., and Tomasello, M. 1999: A nonverbal false belief task: The performance of children and great apes. *Child Development*, 70, 381–395.

Call, J., and Tomasello, M. (in press). What do chimpanzees know about seeing revisited: An explanation of the third kind. In N. Eilan, C. Hoerl, T. McCormack, and J. Roessler (eds.), *Issues in Joint Attention*. Oxford University Press.

Carpenter, M., Akhtar, N., and Tomasello, M. 1998b: Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior & Development*, 21, 315–330.

Carpenter, M., Call, J., and Tomasello, M. (in press, a): Understanding others' prior intentions enables 2-year-olds to imitatively learn a complex task. Child Development.

Carpenter, M., Call, J., and Tomasello, M. (in press, b): Some 36-month-old children understand false beliefs. *British Journal of Developmental Psychology*

Carpenter, M., Nagell, K., and Tomasello, M. 1998a: Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, Volume 255.

Cheney, D.L., and Seyfarth, R.M. 1990: *How monkeys see the world*. Chicago: University of Chicago Press.

Clark, E.V. 1978: Awareness of language: Some evidence from what children say and do. In A. Sinclair, R. Jarvella, and W.J.M. Levelt (eds.), *The child's conception of language*. New York: Springerverlag. Pp. 17–43.

Clark, E.V. 1997: Conceptual perspective and lexical choice in acquisition. *Cognition*, 64(1), 1–37.

Clements, W., and Perner, J. 1994: Implicit understanding of belief. *Cognitive Development*, 9(4), 377–395.

Currie, G. 1998: Pretence, pretending and metarepresentation. *Mind & Language*, 13(1), 35–55.

de Villiers, J.G., and de Villiers, P.A. 2000: Linguistic determinism and the understanding of false beliefs. In P. Mitchell and K.J. Riggs (eds). *Children's reasoning and the mind*. (pp. 191–228). Hove, England: Psychology Press.

Diessel, H., and Tomasello, M. 2001: The acquisition of finite complement clauses in English: A usage based approach to the development of grammatical constructions. *Cognitive Linguistics*, 12, 97–141.

UNCORRECTED PROOF

© *Blackwell Publishing Ltd. 2003*

Dunn, J., Brown, J., and Beardsall, L. 1991: Family talk about feeling states and children's later understanding of others' emotions. *Developmental Psychology*, 27, 448–455.

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

Farrar, J., and Maag, L. 2002: Early language development and the emergence of a theory of mind. *First Language*, 22, 197–214.

Fodor, J.A. 1992: A theory of the child's theory of mind. *Cognition*, 44(3), 283–296.

Gergely, G. 2001: The development of understanding self and agency. In U. Goswami (ed.), *Blackwell's Handbook of Childhood Cognitive Development*. Oxford: Blackwell.

Gergely, G., Bekkering, H., and Király, I. 2002: Rational imitation of goal-directed actions. *Nature*, 415, 755.

Gergely, G., Nadasdy, Z., Csibra, G., and Biro, S. 1995: Taking the intentional stance at 12 months of age. *Cognition*, 56(2), 165–193.

German, T.P., and Johnson, S.C. 2002: Function and the origins of the design stance. *Journal of Cognition and Development*, 3 (3), 279–300.

Gordon, R. 1995: Simulation without introspection or inference from me to you. In M. Davies and T. Stone (eds.), *Mental Simulation: Evaluations and applications*. Oxford: Blackwell.

Gordon, R. 1996: Radical simulationism. In P. Carruthers and P.K. Smith (eds.), *Theories of theories of mind* (pp. 11–21). Cambridge: Cambridge University Press.

Hare, B., Call, J., Agnetta, B., and Tomasello, M. 2000: Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, 59, 771–785.

Hare, B., Call, J., and Tomasello, M. 2001: Do chimpanzees know what conspecifics know? *Animal Behaviour*, 61, 139–151.

Harris, P. 1996: Desires, beliefs and language. In P. Carruthers & P.K. Smith (eds.), *Theories of theories of mind* (pp. 200–220). Cambridge: Cambridge University Press.

Harris, P.L. 1999: Acquiring the art of conversation: Children's developing conception of their conversation partner. In M. Bennett (ed.), *Developmental psychology: Achievements and prospects*. (pp. 89–105). Philadelphia: Psychology Press.

Hobson, P.R. 1993: *Autism and the development of mind*. Hillsdale, NJ: Erlbaum.

Hobson, P.R. 2002: *The cradle of thought*. London: Macmillan.

Horgan, D. 1981: Learning to tell jokes: A case study in metalinguistic awareness. *Journal of Child Language*, 8, 217–224.

Johnson, K.E., and Mervis, C.B. 1997: First steps in the emergence of verbal humor: a case study. *Infant Behaviour and Development*, 20 (2), 187–196.

Karmiloff-Smith, A. 1992: *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.

Kruger, A.C. 1990: Coming to consensus with peers and adults: A microanalysis of problem-solving processes. *Merrill-Palmer Quarterly*.

Kruger, A., and Tomasello, M. 1986: Transactive discussions with peers and adults. *Developmental Psychology*, 22, 681–685.

Kruger, A., and Tomasello, M. 1996: Cultural learning and learning culture. In D. Olson (ed.), *Handbook of Education and Human Development: New Models of Teaching, Learning, and Schooling*. Oxford: Blackwell.

UNCORRECTED PROOF

© *Blackwell Publishing Ltd. 2003*

Leslie, A. 1994: ToMM, ToBY, and agency: Core architecture and domain specificity. In L. Hirschfeld and S. Gelman (eds.), *Mapping the mind*. Cambridge University Press.

Lohmann, H. (in preparation): The role of language in false belief understanding. Doctoral Dissertation.

Lohmann, H., and Tomasello, M. (in press): The Role of Language in the Development of False Belief Understanding: A Training Study. *Child Development*.

Matan, A., and Carey, S. 2001: Changes within the core of artifact concepts. *Cognition*, 77, 1–26

Mead, G. 1934: *Mind, self, and society*. University of Chicago Press.

Meltzoff, A. 1995: Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31, 838–850.

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

Peterson, C.C., and Siegal, M. 2000: Insights into theory of mind from deafness and autism. *Mind & Language*, 15(1), 123–145.

Povinelli, D.J., Bering, J.M., and Giambrone, S. 2000: Toward a science of other minds: Escaping the argument by analogy. *Cognitive Science*, 24, 509–541.

Rakoczy, H., Tomasello, M., and Striano, T. 2002: Children and 'Virgin Objects'—learning pretense and instrumental actions with new things. Submitted for publication.

Saussure, F. 1916: *Cours de linguistic générale*. Lausanne-Paris: Payot.

Searle, J. 1995: *The construction of social reality*. New York: Free Press.

Sigman, M., and Capps, L. 1997: *Children with autism*. Harvard University Press.

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

Striano, T., Tomasello, M., and Rochat, P. 2001: Social and object support for early symbolic play. *Developmental Science*, 4, 442–455.

Tomasello, M. 1990: Cultural transmission in the tool use and communicatory signaling of chimpanzees? In S.T. Parker and K.R. Gibson (eds.), *'Language' and intelligence in monkeys and apes: Comparative developmental perspectives*. (pp. 274–311). New York, NY, USA: Cambridge University Press.

Tomasello, M. 1995: Joint attention as social cognition. In C. Moore and P.J. Dunham (eds.), *Joint attention. Its origin and role in development* (pp. 103–130). Hillsdale, NJ: Lawrence Erlbaum.

Tomasello, M. 1996: Do apes ape? In C.M. Heyes and B.G. Galef, Jr. (eds.), *Social learning in animals: The roots of culture*. (pp. 319–346). NY: Academic Press.

Tomasello, M. 1998: Reference: Intending that others jointly attend. *Pragmatics and Cognition*, 6, 219–234.

Tomasello, M. 1999: *The cultural origins of human cognition*. Harvard University Press.

Tomasello, M., and Call, J. 1997: *Primate cognition*. New York: Oxford University Press.

Tomasello, M., and Camaioni, L. 1997: A comparison of the gestural communication of apes and human infants. *Human Development*, 40(1), 7–24.

Tomasello, M., and Haberl, K. (submitted): Understanding attention: 12- and 18-month-olds know what's new for other persons.

Tomasello, M., Call, J., and Hare, B. 1998: Five primate species follow the visual gaze of conspecifics. *Animal Behaviour*, 55, 1063–69.

UNCORRECTED PROOF

© *Blackwell Publishing Ltd. 2003*

Tomasello, M., Hare, B., and Agnetta, B. 1999: Chimpanzees follow gaze direction geometrically. *Animal Behaviour*, 58, 769–77.

Tomasello, M., Kruger, A., and Ratner, H. 1993: Cultural learning. *Behavioral and Brain Sciences*, 16, 495–552.

von Hofsten, C., and Siddiqui, A. 1993: Using the mother's actions as a reference for object exploration in 6- and 12-month-old infants. *British Journal of Developmental Psychology*, 11, 61–74.

Vygotsky, L. 1978: Mind in society: The development of higher psychological processes, ed. M. Cole. Harvard University Press.

Wellman, H., and Bartsch, K. 1994: Before belief: Children's early psychological theory. In C. Lewis and P. Mitchell (eds.), *Children's early understandiong of mind*. Erlbaum.

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

Woodward, A.L. 1998: Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1), 1–34.

UNCORRECTED PROOF

© *Blackwell Publishing Ltd. 2003*