BEHAVIORAL AND BRAIN SCIENCES (2005) **28,** 675–735
Printed in the United States of America

Understanding and sharing intentions: The origins of cultural cognition

Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll
Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany
tomas@eva.mpg.de carpenter@eva.mpg.de call@eva.mpg.de
behne@eva.mpg.de moll@eva.mpg.de
http://www.eva.mpg.de/psycho/

Abstract: We propose that the crucial difference between human cognition and that of other species is the ability to participate with others in collaborative activities with shared goals and intentions: shared intentionality. Participation in such activities requires not only especially powerful forms of intention reading and cultural learning, but also a unique motivation to share psychological states with others and unique forms of cognitive representation for doing so. The result of participating in these activities is species-unique forms of cultural cognition and evolution, enabling everything from the creation and use of linguistic symbols to the construction of social norms and individual beliefs to the establishment of social institutions. In support of this proposal we argue and present evidence that great apes (and some children with autism) understand the basics of intentional action, but they still do not participate in activities involving joint intentions and attention (shared intentionality). Human children's skills of shared intentionality develop gradually during the first 14 months of life as two ontogenetic pathways intertwine: (1) the general ape line of understanding others as animate, goal-directed, and intentional agents; and (2) a species-unique motivation to share emotions, experience, and activities with other persons. The developmental outcome is children's ability to construct dialogic cognitive representations, which enable them to participate in earnest in the collectivity that is human cognition.

Keywords: collaboration; cooperation; cultural learning; culture; evolutionary psychology; intentions; shared intentionality; social cognition; social learning; theory of mind; joint attention

Human beings are the world's experts at mind reading. As compared with other species, humans are much more skillful at discerning what others are perceiving, intending, desiring, knowing, and believing. Although the pinnacle of mind reading is understanding beliefs – as beliefs are indisputably mental and normative – the foundational skill is understanding intentions. Understanding intentions is foundational because it provides the interpretive matrix for deciding precisely what it is that someone is doing in the first place. Thus, the exact same physical movement may be seen as giving an object, sharing it, loaning it, moving it, getting rid of it, returning it, trading it, selling it, and on and on – depending on the goals and intentions of the actor. And whereas understanding beliefs does not emerge until around age 4 in human ontogeny, understanding intentions begins to emerge at around a child's first birthday.

Human beings are also the world's experts at culture. Humans do not just interact with conspecifics socially, as do many animal species, but they also engage with them in complex collaborative activities such as making a tool together, preparing a meal together, building a shelter together, playing a cooperative game, collaborating scientifically, and on and on. These collective activities and practices are often structured by shared symbolic artifacts, such as linguistic symbols and social institutions, facilitating their "transmission" across generations in ways that ratchet them up in complexity over historical time. Children become more skillful at collaborating and interacting with others culturally throughout early childhood, but their first nascent attempts begin, once again, at around the first birthday.

Tomasello et al. (1993) argued and presented evidence that these two dimensions of human expertise – reading intentions and interacting with others culturally – are intimately related. Specifically, the way humans understand the intentional actions and perceptions of others creates species-unique forms of cultural learning and engagement, which then lead to species-unique processes of cultural cognition and evolution. For example, it is only if a young child understands other persons as intentional agents that she can acquire and use linguistic symbols – because the learning and use of symbols requires an understanding that the partner can voluntarily direct actions and attention to outside entities. Indeed, material and symbolic artifacts of all kinds, including even complex social institutions, are in an important sense intentionally constituted (Bloom 1996; Searle 1995; Tomasello 1999a).

Recently, however, some new empirical findings have emerged which suggest that understanding intentions cannot be the whole story of cultural cognition. Briefly, the main finding is that some nonhuman primates understand more about intentional action and perceptions than was previously believed (and this is also true, to some degree, of children with autism). But they do not thereby engage socially and culturally with others in the ways that human children do. Therefore, understanding the intentional actions and perceptions of others is not by itself sufficient to produce humanlike social and cultural activities. Something additional is required.

Our hypothesis for this "something additional" is shared intentionality. We propose that human beings, and only human beings, are biologically adapted for participating in collaborative activities involving shared goals and socially coordinated action plans (joint intentions). Interactions of this type require not only an understanding of the goals, intentions, and perceptions of other persons, but also, in addition, a motivation to share these things in interaction with others – and perhaps special forms of dialogic cognitive representation for doing so. The motivations and skills for participating in this kind of "we" intentionality are woven into the earliest stages of human ontogeny and underlie young children's developing ability to participate in the collectivity that is human cognition.

In this article, we explicate and elaborate this account of how humans come to (1) understand intentional action and (2) participate in activities involving shared intentionality. Our focus is on how these two skills interweave during normal human ontogeny, but we also review recent empirical findings with great apes and children with autism, providing the skeleton of an evolutionary account in the process. We employ a "control systems" approach (from cybernetic theory) to characterize the structure of intentional action and a "shared intentionality" approach (from the philosophy of action) to characterize the types of cognitive skills and social engagements that make possible uniquely human activities such as the creation and use of linguistic and mathematical symbols, the creation and use of artifacts and technologies that accumulate modifications over generations in cultural evolution, and the creation of social practices and institutions such as marriage and government that depend on collective beliefs – in short, what we will call *skills of cultural cognition*.

Michael Tomasello is Co-Director of the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. Research interests focus on processes of social cognition, social learning, and communication/language in human children and great apes.

Malinda Carpenter is a Staff Scientist in the Department of Developmental and Comparative Psychology at the Max Planck Institute for Evolutionary Anthropology. Her research interests include imitation, gestural communication, and understanding of others' intentions in typically developing children, children with autism, and apes.

Josep Call is Co-director of the Wolfgang Koehler Primate Research Center and Senior Scientist at the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. His research interests focus on problem solving, causal understanding, social cognition, and social behavior in primates and other animals.

Tanya Behne is a postdoctoral associate at the Max Planck Institute for Evolutionary Anthropology. Her current research focuses on children's social cognitive development, especially infants' understanding of intentional action, imitative learning, and understanding communicative intent.

Henrike Moll is a Ph.D. student in the Department of Developmental and Comparative Psychology at the Max Planck Institute for Evolutionary Anthropology. Research interests focus on human children's and infants' understanding of perception and attention.

1. Intentional action

If we want to know how people understand intentional action, we must first have a model of exactly what intentional action is. Here we propose a simple model based on control-systems principles – in which goal, action, and perceptual monitoring are all seen as components in the larger adaptive system that serves to regulate the organism's behavioral interactions with the environment.

As discovered by cyberneticians such as Weiner (1948) and Ashby (1956), machines that act on their own "intelligently" all have the same basic organization involving the same three components: (1) a reference value or *goal* toward which the system acts, (2) the ability to act in order to change the environment, and (3) the ability to perceive the environment so as to know when the state of the environment matches the reference value. The prototypical exemplar, of course, is the thermostat which – all by itself without human intervention – can regulate the temperature of a room. It does this by (1) having a reference value set by a human (e.g., 25 degrees), (2) being able to turn on or off an air heater or cooler, and (3) being able to sense the room temperature (e.g., with a thermometer) and compare it to the reference value to determine whether heating, cooling, or no action is required. This circular organization – goal determines action, which changes perception (feedback), which (when compared to goal) again determines action – makes the thermostat a self-regulating device.

The application of this insight to human intentional action is depicted in Figure 1, using the example of an individual faced with a closed box and wanting it open. This diagram embodies a number of the terminological conventions we will use in our review of the empirical literature, as well as some substantive points about how we think intentional action works. To begin at the top of the figure, the word *goal* contains a systematic ambiguity that has contributed to much confusion (e.g., see Want & Harris 2001). When it is said that a person wants a box open, for example, we may distinguish the external goal – a certain state of the environment such as an open box – and the internal goal – an internal entity that guides the person's behavior (e.g., a mental representation of a desired state such as an open box). We will reserve the term goal for the internal goal, and for the external goal we will use such expressions as "the desired result".

Another important distinction that is not always clearly made is that between goal and intention. Following Bratman (1989), we propose that an intention is a plan of action the organism chooses and commits itself to in pursuit of a goal. An intention thus includes both a means (action plan) as well as a goal (in Fig. 1, the intention includes both the goal of an open box as well as the action plan chosen to make that happen). The fact that the intention includes the goal

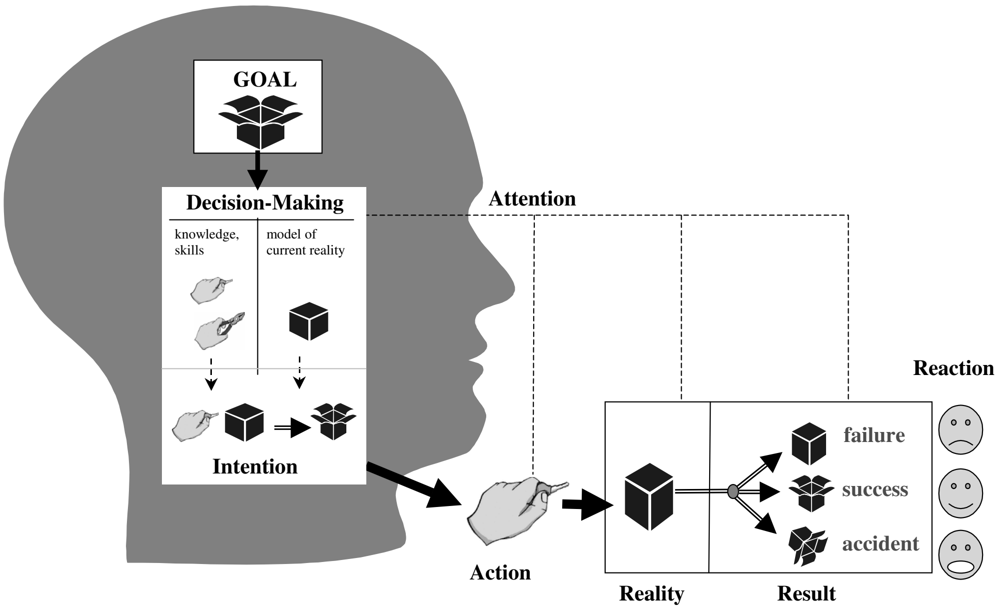

Figure 1. Human intentional action. The goal is an open box; reality is a closed box. The actor chooses a means (plan), depicted as hands doing things, which forms an intention. The resulting action causes a result, which leads to a reaction from the actor.

explains why the exact same action may be considered different things intentionally; for example, cutting the box as an act may be either "opening it" or "making kindling," depending on the goal. So the organism has the goal "that X be the case" and the intention "to do A" in pursuit of that goal. In choosing an intended course of action (*decision making* in Fig. 1), the organism consults both its stored knowledge/skills and its mental model of current reality – that is, those aspects that are "relevant" to the goal. The chosen action is "rational" to the degree that it effectively accommodates the organism's knowledge, skills, and model of current reality.

Moving out of the organism and into the realm of what is observable from outside, the organism's intention typically results in concrete behavioral action of one sort or another (*large hand* in Fig. 1). This is often accompanied by such things as signs of effort and direction of gaze. Also relevant is current reality – a *closed box* in Figure 1 – and any additional constraints in the context (e.g., a lock on the box). After the action on reality has taken place, the state of the world is transformed in one way or another (including no change), and we call this the result of the action, which is also typically observable. In Figure 1, we can see various ways that the result may or may not match the goal: (1) a failed attempt, in which the action does not succeed in changing the state of reality to meet the goal; (2) success in which the action changes reality so as to match the goal; and (3) an accident, which is also not successful but for different reasons (the action causes an unintended result). Quite often, each of these results is accompanied by an emotional reaction on the part of the behaving organism: disappointment at failure, happiness at success, and surprise at an accident (also depicted in Fig. 1). The two types of results representing failure are typically followed by persistent, often variable, efforts toward the goal.

Finally, crucial to the whole process is the organism's perceptual monitoring throughout (the *dashed lines* in Fig. 1). The organism monitors the situation to see (1) what is the current reality (information it uses continuously), (2) whether it executed the action intended, and (3) the result produced by the action. In Figure 1, the label used is not perception but attention. The reason is that in each of these cases the organism is not perceiving everything, but rather it is attending to just those aspects of the situation that are relevant to the goal at hand. Thus, the organism may not pay attention to the color of the box, the temperature of the room, or other things unrelated to its goal. As we have argued previously (e.g., see Tomasello 1995), attention may thus be thought of as intentional perception (selective attention). This monitoring process thus completes the circular arrangement characteristic of intentional action: the organism acts so as to bring reality (as it perceives it) into line with its goals.[1]

Two complications. First, it is important to recognize the hierarchical structure involved here (Powers 1973). Once the organism chooses an action plan to enact in intentional action, it typically must also create lower-level goals and action plans. For example, in Figure 1 the plan chosen for achieving the goal of an open box might involve opening it with a key. This requires having an appropriate key in hand (as subgoal), which means creating a subplan to walk to the nearby drawer, open it, fetch the key, return to the box, and use the key. At each step of choosing a subgoal and subplan, there are potentially multiple possibilities to choose from, and these must be assessed with respect to their predicted efficacy – what we will call *decision making*. And we must not forget the higher-level goals either. The organism wants the box open for a reason; perhaps it has a higher-level goal of obtaining the birthday gift sent by Uncle Ralph, and therefore opening the box is, from this higher perspective, only a means. In general, what is a goal when viewed from beneath is a means when viewed from above. Starting at any given level, moving up to more general goals explains *why* a person has a particular goal: she wants the box open in order to obtain the gift. Moving down the hierarchy to more specific action plans specifies *how* a goal is achieved in terms of intentional actions: she intends to open the box by using a key.

Second, a related complication is that an organism may have as a goal some movement or action in itself; for example, a dancer's goal is simply to perform certain body movements that have no observable environmental effects. And it may also happen that an object-related goal includes as a component a specific action. Thus, as a child approaches the box, we might think either that her goal is that it be open (and the means chosen to do it is cutting with scissors) or, alternatively, that her goal is that she open it by cutting with her new scissors. The distinguishing test is easy. If we open the box before the child arrives, in the first case she will be happy (she only wanted it open), whereas in the second case she will be unhappy (she wanted to do it herself by cutting with her new scissors). This complication – that organisms may have as goals either environmental effects or self-actions or some combination of both – plays a crucial role in imitation, since the imitator often must decide whether to do something effectively or else in the way a demonstrator has done it. It also plays a role in some collaborative activities in which the goal is not just that something be done but that it be done together with someone else. Basically, the state of the world the organism seeks to bring about – its goal – may include just about anything in particular cases, including self-action and joint action with others.

This is our model of intentional action. But our concern is not with the question of whether organisms themselves produce intentional actions, which many do, but rather it is with the question of how they understand the intentional actions of others. Our special concern is with human ontogeny and when and how this understanding emerges.

2. Understanding intentional action

The classic studies of children's understanding of intentions are studies in which adults ask preschool children explicit verbal questions about various kinds of actions – for example, successful, accidental, and unsuccessful – and they respond verbally. For example, Piaget (1932) presented children with stories in which a child did things either "on purpose" or "by accident" and asked about blameworthiness and the like. In other studies, children observe actions and then are asked specific questions about the goals and intentions of the actors (e.g., see Baird & Moses 2001; Smith 1978; Shultz & Wells 1985). Recently, the focus has been on whether children distinguish desires (or goals) from intentions (or plans), and the general finding is that they can do so in their explicit language from about 5 years of age (e.g., see Feinfeld et al. 1999; Schult 2002). Also interesting are studies in which preschool-age children talk about artifacts and artwork in terms of the intentions of those who produced them (e.g., see Bloom & Markson 1998; Gelman & Ebeling 1998).

But children actually begin to demonstrate an understanding of intentional action long before this, during infancy, and our primary concern is with these ontogenetic origins. Even in the first year or so of life, we may distinguish three levels in children's understandings of the actions of others (here and throughout the observer is *she* and the actor is *he*).

Acting animately. An observer perceives that the actor has generated his motion autonomously; that is, she distinguishes animate *self-produced* action from inanimate, caused motion. There is no understanding that the actor has a goal, and so means and ends are not distinguished, nor are successful and unsuccessful actions. Although observers may learn from experience what animate actors typically do in familiar situations, predicting behavior in novel circumstances is basically impossible. (In the format of Fig. 1, inside the actor's head is nothing.)

Pursuing goals. An observer perceives and understands that the actor has a goal and behaves with *persistence* until reality matches the goal; that is, she understands that the actor recognizes the success or failure of his actions with respect to the goal and continues to act in the face of failure. This understanding implies that the observer also knows that the actor *sees* things (e.g., objects with respect to which he has goals, potential obstacles to goals, the results of actions) and that this helps to guide action and determine satisfaction with results. Understanding action in this way enables observers to predict what actors will do in at least some novel situations. (In the format of Fig. 1, inside the actor's head is a goal and perceptual monitoring.)

Choosing plans. An observer perceives and understands that the actor considers action plans and *chooses* which of them to enact in intentional action (and these may be more or less rational depending on their fit with perceived reality). She also understands that in acting toward a goal the actor chooses which entities in its perceptual field to *attend to*. In general, the observer understands that actors act and attend to things for reasons, which enables her to predict what an actor will do in a wide variety of novel situations. (All elements of Fig. 1 present.)
Children's understanding of these different aspects of intentional action and perception emerge, in this order, at different points in infancy.

2.1. Understanding animate action

Infants recognize self-produced, biological motion within a few months after birth (Bertenthal 1996), and they soon turn to look in the same direction as other persons as well (D'Entremont et al. 1997). By around 6 months of age, infants have developed sufficient expectations about human animate action to be able to predict what others will do in familiar situations. Thus, for example, using an habituation methodology, Woodward (1998) found that infants of this age expect people (specifically, human hands) to do such things as reach for objects they were just reaching for previously. Infants do not expect inanimate objects that resemble human hands (e.g., a garden-tool "claw") to "reach" toward the familiar object in similar circumstances.

This and similar studies are sometimes interpreted as demonstrating that 6-month-olds see human actions as goal directed (e.g., see Woodward 1999). From our perspective, a more felicitous appellation would be object directed; that is, infants in these studies clearly expect the adult to be consistent in his interactions with the same object over a short span of time, and they follow gaze to the object he is looking at. But to do these things, infants need only to understand that people spontaneously produce behavior (they are animate beings) and to have some familiarity with what people typically do in familiar circumstances; they do not need to have any understanding of the internal structure of intentional actions. For example, they do not need to know that the actor is evaluating the efficacy of his action toward a goal and persisting in his behavior until he is successful – much less that he chooses an action to enact intentionally for "rational" reasons.

2.2. Understanding the pursuit of goals

By 10 months of age, infants segment streams of continuous behavior into units that correspond to what adults would see as separate goal-directed acts (Baldwin et al. 2001). Infants of this same age also look to an adult's face when he teases her with a toy or obstructs her play with a toy (Carpenter et al. 1998b; Phillips et al. 1992) – perhaps suggesting that infants are seeking information about the adult's goal by trying to discern where he is looking or his emotional state.

But more than segmenting actions and trying to identify goals, infants of this age also demonstrate an ability to understand an actor's persistence to a goal – which involves an understanding that actors perceptually monitor and recognize when their actions have changed the world in the desired way. This is clearest in the case of actions that are not immediately successful, because in this case the child must infer the actor's goal even though it is not achieved (and therefore not observed) from various aspects of behavior and context. The two main categories of unsuccessful actions are trying and accidents.

First, infants' understanding of trying is evident in the well-known series of habituation studies by Gergely and colleagues involving obstacles (Csibra et al. 1999, 2002; Gergely et al. 1995). In the classic study, infants were habituated to a large dot "jumping" over an obstacle and approaching a small dot. Later, with the obstacle gone, 9- and 12-month-olds (but not 6-month-olds) dishabituated to the same jumping motion (even though its path of movement was identical to that during habituation), and they did not dishabituate to the large dot going directly to the small dot (even though this was a new motion). The argument is that infants remained habituated to the different motion in this latter condition because they saw the large dot's actions as in some sense the same as during habituation: goal-directed and efficient action to the small dot. It thus seems that 9- to 12-month-old infants understand at least one aspect of trying: actors routinely go around obstacles to get to goals.

In a more interactive methodology, Behne et al. (2005) engaged infants in a game in which an adult gave them toys across a table. Interspersed were trials in which the adult held up a toy but did not give it over. In some cases this was because he was unwilling, in various ways, and in other cases it was because he was trying but unable, in various ways (e.g., could not extract it from a container). In reaction to these activities, 9- to 18-month-olds, but not 6-month-olds, showed more signs of impatience (e.g., reaching, turning away) when the adult kept the toy for himself than when he was making a good faith effort to give it over. Infants thus seemed to have appreciated that in the unable scenarios the adult was, for example, trying to give them a toy as he struggled unsuccessfully against the recalcitrant container. Interestingly, 15-month-old and older infants can even imagine the specific goal an actor is trying to attain as he struggles unsuccessfully – as evidenced by the fact that when they observe unsuccessful actions they imitate not those specific movements but rather they make attempts to reproduce the actor's desired result in the environment using novel actions (Bellagamba & Tomasello 1999; Johnson et al. 2001; Meltzoff 1995).

The second way that infants display an understanding of the persistent nature of goal-directed activity is when they distinguish purposeful actions from accidental actions, knowing that an accidental action will not satisfy the actor's goal. Thus, in the Behne et al. study, another pair of conditions involved an adult either holding out a toy in a teasing fashion (unwilling) or holding out a toy but dropping it accidentally (unable). In reaction to these two different adult goals, 9-month-old (and older) infants, but not 6-month-old infants, were more impatient when the adult was teasing them than when he was simply being clumsy. The earliest age at which children first understand accidental actions thus matches the age at which they first understand trying actions (as determined by two different experimental paradigms): 9 months, but not 6 months. Relatedly, Carpenter et al. (1998a) found that 14- to 18-month-old infants chose to imitate purposeful but not accidental actions.

When 9-month-olds begin to understand that actors are pursuing goals, they must know also that the actor perceives his actions and their results. Only if infants understand this can they understand why the actor is satisfied or disappointed after completing an action. So in addition to 6-month-olds' gaze following, it is important that 12-month-olds (younger infants have not been tested) follow the direction of adult gaze in more complex situations, for example, to locations behind barriers (Moll & Tomasello 2004). This behavior goes beyond simple gaze following, because the infant does not just respond to a head turn by turning her own head in the same direction, but she actually has to locomote some distance to attain the appropriate viewing angle – indicating an understanding that the adult sees something that she does not (see also Caron et al. 2002 for studies in which infants in this same age range know that the adult's visual access is impeded by barriers).

A reasonable conclusion from all of this is thus that 9- to 12-month-old infants understand the basics of goal-directed action. They understand that actors try to achieve goals, that they keep trying persistently after failed attempts and accidents and around obstacles, and that when they succeed they stop acting toward the goal – which involves an understanding that people perceptually monitor their actions so that they can recognize when they have succeeded. But this is still not all that can be known about intentional action.

2.3. Understanding the choice of plans

In the months immediately following their first birthdays, infants begin to understand that, in pursuing a goal, an actor may consider various action plans (means) and chooses one to enact in intentional action based on some reason related to reality. There is only one study demonstrating such understanding in young infants. It involves so-called rational imitation.

Gergely et al. (2002) showed 14-month-old infants an adult touching his head to the top of a box to turn on a light. However, for half of the infants, the adult's hands were occupied during this action (he was shivering and holding a blanket around his shoulders) and, for the other half, the adult's hands were free during the action. In both conditions, infants thus saw that the adult was trying to turn on the light with his head. Nevertheless, when it was their turn (and they had no blanket around their shoulders), infants who saw the hands-free demonstration bent over and touched the box with their heads more often than did infants who saw the hands-occupied demonstration. Apparently, infants assumed that if the adult's hands were free and he still chose to use his head, then there must be a good reason for this choice – he intended to turn on the light with his head – and so they followed suit. However, if the adult's hands were occupied, then the use of the head was explained away as necessary given his circumstance – without the constraint of the blanket he would not have chosen this means – and so they were free to ignore it since the same constraint was not present for them. In this study, therefore, infants understood not just that the actor perceived and evaluated the efficacy of his actions to a goal, but rather infants understood that the actor perceived and evaluated reality rationally before choosing an action plan designed to accommodate this reality in pursuit of the goal.[2]

In terms of the understanding of perception, infants at this age seem to have an understanding of at least some aspects of selective attention. Tomasello and Haberl (2003) had an adult say to 12- and 18-month-old infants "Oh, wow! That's so cool! Can you give it to me?" while gesturing ambiguously in the direction of three objects. Two of these objects were "old" for the adult – he and the child had played together with them – and one was "new" to him (though not to the child). Infants gave the adult the object that was new for him. This suggests that they understood that even though the adult was looking at and seeing all three objects equally, he was selectively attending only to the one that he had not previously experienced and so now wanted. One interpretation of this result is that infants understand perception as a kind of rational action also, in the sense that from all the things they see people choose to attend to only a subset, and they do this for reasons related to their goals.

2.4. Cultural learning

The developmental picture that emerges is thus as follows. Six-month-old infants perceive animate action and follow gaze direction, which enables them to build up experiences on the basis of which they predict people's actions in familiar contexts. By 9 months of age, infants understand that that people have goals and persist in behaving until they see that their goal has been reached (avoiding obstacles and persisting past accidents and failures in the process) – being happy when the goal is reached and disappointed if it is not. By 14 months of age, infants begin to understand full-fledged intentional action – including the rudiments of the way people make rational decisions in choosing action plans for accomplishing their goals in particular reality contexts and selectively attending to goal-relevant aspects of the situation.

This kind of understanding leads to some powerful forms of cultural learning, especially imitative learning in which the observer must perform a means-ends analysis of the actor's behavior and say in effect "When I have the same goal I can use the same means (action plan)." This analysis is also necessary before one can ask why someone did something and whether that reason also applies in my circumstance ("rational imitation"). Without such an analysis, only simpler forms of social learning are possible (Tomasello et al. 1993, and see sect. 4.1.1). The main point is that 1-year-old infants use their newly emerging skills of intention understanding not only to predict what others will do, but also to learn from them how to do things conventionally in their culture.

3. Shared intentionality

When individuals who understand one another as intentional agents interact socially, one or another form of shared intentionality may potentially emerge. Shared intentionality, sometimes called "we" intentionality, refers to collaborative interactions in which participants have a shared goal (shared commitment) and coordinated action roles for pursuing that shared goal (Gilbert 1989; Searle 1995; Tuomela 1995). The activity itself may be complex (e.g., building a building, playing a symphony) or simple (e.g., taking a walk together, engaging in conversation), so long as the interactants are engaged with one another in a particular way. Specifically, the goals and intentions of each interactant must include as content something of the goals and intentions of the other. When individuals in complex social groups share intentions with one another repeatedly in particular interactive contexts, the result is habitual social practices and beliefs that sometimes create what Searle (1995) calls social or institutional facts: such things as marriage, money, and government, which only exist due to the shared practices and beliefs of a group.

According to Bratman (1992), joint cooperative activities, as he calls them, have three essential characteristics that distinguish them from social interaction in general (here modified slightly): (1) the interactants are mutually responsive to one another, (2) there is a shared goal in the sense that each participant has the goal that we (in mutual knowledge) do X together, and (3) the participants coordinate their plans of action and intentions some way down the hierarchy – which requires that both participants understand both roles of the interaction (*role reversal*) and so can at least potentially help the other with his role if needed. Some aspects of this account of shared intentionality are translated into our diagrammatic conventions in Figure 2.

Note two things about Figure 2, which is meant to depict each participant's understanding of the interaction. First and most important, the cognitive representation of the goal contains both self and other; that is, it contains not only the self's goal that the box be open, but also the self's goal that this be accomplished with the partner. One might simply say, then, that his goal concerns their mutual actions. But since he does not have expectations about the partner's particular behaviors, but rather about her intentional actions (as defined by goals such as opening the box), we may better say that the actor wants his interactant to have, along with him, the goal of opening the box – which she should pursue using whatever means are necessary. And of course the partner, assuming she also desires collaboration, also wants her partner to share her goal – thus creating a "shared commitment" (Gilbert 1989). And so, overall, this figure instantiates our claim that there is a special kind of shared motivation in truly collaborative activities in the form of a shared goal – each interactant has goals with respect to the other's goals – a crucial point to which we return later in differentiating human collaboration and intentional communication from the social interactions of other primate species.[3]

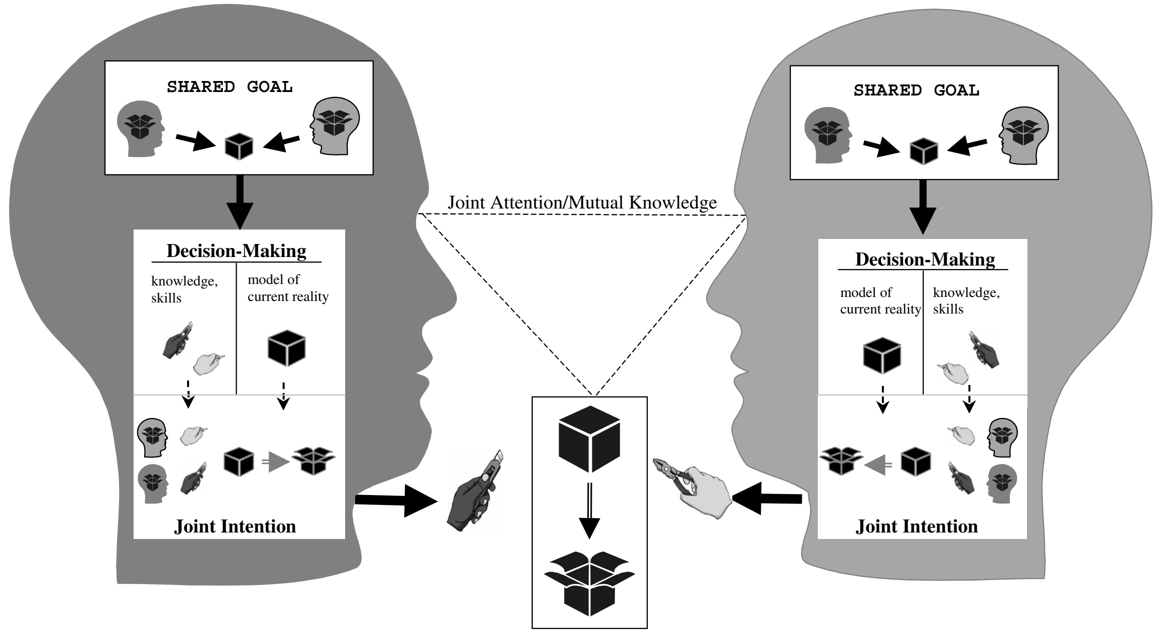

Figure 2. Each partner's conception of a collaborative activity in which a shared goal and joint intention (with complementary roles) are formed.

The second important aspect of this figure is that the cognitive representation of the intention also contains both self and other – it is thus a joint intention. This is necessary because both collaborators must choose their own action plan in the activity in light of (and coordinated with) the other's action plan: my role is to hold the box steady while you cut it open. This requires that each participant cognitively represent both roles of the collaboration in a single representational format – holistically, from a "bird's-eye view," as it were – thus enabling role reversal and mutual helping. Overall, then, collaborative activities require both an alignment of self with other in order to form the shared goal, and also a differentiation of self from other in order to understand and coordinate the differing but complementary roles in the joint intention.

In the first year or so of life, human infants socially interact with other persons in various ways leading gradually to more or less full participation in activities involving shared intentionality.

Dyadic engagement: Sharing behavior and emotions. An individual interacts with, and is mutually responsive to, an animate agent directly – mainly through the expression of emotions and behavioral turn taking. (In the format of Fig. 2, nothing inside the heads.)

Triadic engagement: Sharing goals and perception. An individual interacts together with a goal-directed agent toward some shared goal. In doing this, both interactants perceptually monitor the goal-directed behavior and perceptions of the partner. (In the format of Fig. 2, inside the heads are shared goals and perceptual monitoring.)

Collaborative engagement: Joint intentions and attention. An individual interacts with an intentional agent toward some shared goal and with coordinated action plans as manifest in a joint intention – and with joint attention (mutual knowledge) as well. Each interactant thus cognitively represents both the shared goal and action plans involving complementary roles – with the possibility of reversing roles and/or helping the other in his role, if necessary. (In the format of Fig. 2, all components present.)

These different types of social engagement – which emerge in human ontogeny in this order – depend on particular ways of understanding intentional action in general: as animate, goal directed, or intentional, as elaborated in the previous section. In addition, however, they also rely – in a way to be explained now – on a special motivation to share psychological states with other persons.

3.1. Dyadic engagement: Sharing behavior and emotions

Human infants are extremely sensitive to social contingencies. In their face-to-face interactions with adults, infants from just a few months of age display the ability to take turns in the sense of acting when the adult is more passive and being more passive when the adult is acting (Trevarthen 1979). When these contingencies are broken – for example, in experiments in which the adult's behavior is preprogrammed (or played to the infant over delayed video) – infants show various signs of being out of sorts (for reviews, see Gergely & Watson 1999 and Rochat & Striano 1999). Infants' early social interactions thus clearly show mutual responsiveness on the behavioral level.

But there is another dimension to these interactions that goes beyond simple timing and contingency. Human infants and adults interact with one another dyadically in what are called *protoconversations*. These are social interactions in which the adult and infant look, touch, smile, and vocalize toward each other in turn-taking sequences. But as most observers of infants have noted, the glue that holds protoconversations together is not just contingency but the exchange of emotions (Hobson 2002; Trevarthen 1979). Evidence for this comes from Stern (1985), who found that during protoconversations adult and infant do not just mimic each other or respond randomly, but often express the same emotion using a different behavior (e.g., the adult expresses happiness facially and the child vocally). During protoconversations, infants gaze into the eyes of the partner face-to-face in what is called *mutual gazing*. It is a dyadic activity in the sense that the infant is not monitoring the adult's looking at her or any other object; it is direct engagement.

Although there may be differences in the way protoconversations take place in different cultures – especially in the nature and amount of face-to-face visual engagement – in one form or another they seem to be a universal feature of adult-infant interaction in the human species (Keller et al. 1988; Trevarthen 1993). Protoconversations require not only that the two interactants understand each other as animate agents, but also that they have a special motivation and capacity to share emotions with each other. This additional factor is clearly necessary, as the individuals of many nonhuman species appreciate others as animate agents, but are still unmotivated to engage with them in protoconversations (see sect. 4.1.2 on great apes). But sharing emotions in early infancy is just the beginning of a much longer developmental process. Important though they may be as a foundation, protoconversations do not involve joint commitments to any shared goals or action plans.

3.2. Triadic engagement: Sharing goals and perception

At around 9 to 12 months of age, as infants are beginning to understand other persons as goal directed, they also begin to engage with them in activities that are triadic in the sense that they involve child, adult, and some outside entity toward which they both direct their actions. These are activities such as giving and taking objects, rolling a ball back and forth, building a block tower together, putting away toys together, "pretend" games of eating or drinking, "reading" books, and pointing-and-naming games (Hay 1979; Hay & Murray 1982; Verba 1994). During these activities, infants' looking becomes coordinated with that of the other person triadically toward the relevant outside objects as well. When researchers focus on this aspect of the joint activity, it is most often called "joint attention" (e.g., see papers in Moore & Dunham 1995) – what we will call at this level joint perception.

The question from the point of view of shared intentionality is how the infant understands her engagement with the adult while participating in these initial triadic activities. For instance, suppose a child and adult are building a block tower together. Possibly the child just ignores the adult and places her blocks on the tower irrespective of what the adult is doing; this is not triadic but individual activity. Or perhaps the child is only responsive to the adult in the sense of taking turns; there is no shared goal but only mutual responsiveness. But perhaps adult and child have created a shared goal to build the tower together. This shared goal serves to coordinate their activities around the same object triadically and thereby to enable each participant to know something about what the other is perceiving and to predict what she will do next. The interaction is thus more than sharing behavior or emotions dyadically; it is sharing goals and perceptions with respect to some external entity triadically. Although the evidence is less than fully compelling, Ross and Lollis (1987; see also Ratner & Bruner 1978) observed that, starting at around 9 months of age, infants do a number of things to attempt to reengage a recalcitrant adult in joint activities – such things as handing him an object or gesturing to him to show continued interest in playing the joint game – perhaps suggesting a goal to engage in the activity together (*shared goal*).

Thus, at 9 months of age, infants' special motivation to feel and act and perceive together with others takes on a new form. As infants begin to understand other persons as pursuing goals, their "doing together" with them becomes truly triadic, and the two of them begin to actually share goals as they act together to change the state of the world in some way and to perceive the world together in acts of joint perception. Although nonhuman animals may engage with one another in complex social interactions in which they know the goals of one another and exploit this, they are not motivated to create shared goals to which they are jointly committed in the same way as humans (such that they would be upset if the other reneged; see sect. 4.1.2 on apes). But once again, this is not all that human infants do; there is still further development. Triadic engagements with shared goals still do not necessarily require infants to plan together with others or to coordinate with them the specific intentional actions that will serve as complementary roles in their collaboration.

3.3. Collaborative engagement: Joint intentions and attention

At around 12 to 15 months of age, infants' triadic engagements with others undergo a significant qualitative change. In a classic longitudinal study, Bakeman and Adamson (1984) categorized infants' interactions with their mothers as involving, among other things, either "passive joint engagement" or "coordinated joint engagement." Passive joint engagement referred to triadic interactions in general, whereas coordinated joint engagement referred to triadic interactions in which the infant was much more active in the interaction – not just following adult leads, but also sometimes directing adult behavior and attention as well in a more balanced manner. The empirical finding was that although 9-month-old infants engaged in much passive joint engagement, it was not until 12 to 15 months of age that infants engaged in significant amounts of coordinated joint engagement.

One possible explanation for this change is that soon after their first birthdays infants begin to understand the specific action plans of other persons and something of how they are chosen (as outlined in sect. 1), and they use this understanding in their triadic activities with them. This means, for instance, that the child understands that in pursuing the shared goal of building a block tower the adult holds the edifice steady while she, the child, places blocks. Infants of this age not only share goals but also coordinate roles.

Potential evidence for this interpretation is again provided by Ross and Lollis (1987), who observed that when an adult stopped participating in shared activities, from about 14 months of age infants not only prompted him to reengage, but they sometimes even performed the recalcitrant adult's turn for him. This might suggest that infants of this age understand not only the shared goal but also the two roles involved, and they are motivated to help the adult in his role. Also relevant is the experimental study of Carpenter et al. (2005). They set up situations in which an adult did things like hold out a basket in which the child should place a toy. After the child complied, the adult then placed the basket in front of the child and held the toy himself. Some 12-month-olds, and even more 18-month-olds, then took their turn by holding out the basket for the adult and, importantly, looked to him in anticipation of his placing something in it. It thus seems that after an initial encounter in one role of an interaction, infants often understand the other role – an exchanging of roles that may be called *role-reversal imitation* (see also Ratner & Bruner 1978). One possible explanation for the qualitative shift in infants' social engagements soon after the first birthday, then, is that they are in the process of developing a deeper understanding of intentional action in terms of underlying plans and intentions, and their motivation to share then leads them to create with others not only shared goals but also joint intentions with coordinated roles.

In these interactions, infants are of course also coordinating their perceptions with others, what we will call at this stage joint attention – indicating that infants know that others choose what to attend to within their perceptual fields (as evidenced, for example, by the study by Tomasello & Haberl 2003; see sect. 2.3). It is also at around this same age that infants make their first nascent attempts to establish joint attention actively with others through gestures such as pointing. Of special interest, of course, is declarative pointing in which infants direct adults' attention seemingly for the sole motive of sharing attention. Thus, when an adult reacts to the pointing of a 12-month-old by simply looking to the indicated object, or by looking to the infant (emoting positively), or by doing nothing, infants are not satisfied – implying that these were not their goal. But when an adult responds by looking back and forth from the object to the infant and comments positively, infants are satisfied – implying that this sharing of attention and interest was their goal (Liszkowski et al. 2004). Infants of this age will also sometimes point simply to inform adults of things, even though they themselves have no direct interest in them – a kind of helping motive (Liszkowski et al., in press; see also Kuhlmeier et al. 2003, who found that 12-month-olds discriminate actions in which one computer-animated dot either "helps" or "hinders" another one up an incline). One-year-olds thus seem to have as goals both joint attention itself and also helping others to attain their goals by directing their attention in relevant ways.

Many of these new aspects of triadic interactions come together in a major new accomplishment of children soon after their first birthdays: language. Language, in the sense of linguistic communication, typically begins in earnest at around 13 to 14 months of age. In some theoretical perspectives, language is itself an inherently collaborative activity (Clark 1996) – in at least two senses. First, linguistic symbols are inherently collaborative; they are bidirectional coordination devices, comprising the two implicit roles of speaker and listener. In learning to use symbols, children learn to play both roles and to comprehend both roles no matter which they are playing. Learning symbols thus involves role-reversal imitation (using symbols toward others the way they have used them toward you), and it also involves taking shared perspectives on things and learning that people can choose to attend to things and construe them in many different ways as needed (Clark 1997; Tomasello 1999b).

Second, conversation is an inherently collaborative activity in which the joint goal is to reorient the listener's intentions and attention so that they align with those of the speaker, and joint intentions serve to do that through various kinds of collaborative acts. For example, the speaker collaborates by expressing his communicative intentions in ways that are potentially comprehensible by the listener, even clarifying (helping) when necessary; and the listener collaborates by making good-faith attempts at comprehension by following the speaker's attention-directing signals, making appropriate and relevant inferences, and asking for clarification (help) when needed. Importantly, from their earliest forays into linguistic communication, infants engage in a "negotiation of meaning" in which they request clarification from the adult and produce communicative repairs for the adult when needed (Golinkoff 1993). All of this takes place and is socially structured within the common cognitive ground of various kinds of joint attentional formats (Bruner 1983; Tomasello 1999b) – which make some aspects of entities in the shared situation "mutually manifest" and so potentially "relevant" for acts of interpersonal communication (Sperber & Wilson 1986)

By 12 to 14 months of age, then, the triadic interactions of child and adult around external entities appear as more "coordinated joint engagement," since the child can do such things as reverse roles and help the adult in her role if needed – both necessary for engaging in joint actions embodying joint intentions. In beginning to acquire linguistic symbols at this age, infants again demonstrate an understanding of the different but complementary roles in a social interaction – in this case an interaction involving the exchange of communicative intentions embodied in conventionalized actions – and they are motivated simply to share experience with others and help them toward their goals.

3.4. Cultural creation

And so human infants seem to have from very early in ontogeny a very strong motivation to share emotional states with others, and before the first birthday they express motivations for sharing goals and perceptions with others. By about 12 to 14 months of age, the motivation to share with others reaches down past the sharing of goals and perceptions and into the infant's and others' chosen plans of action and attention: they form joint intentions and participate in joint attention. This means that the child and adult not only construct a shared goal, but they also establish mutually supportive roles by coordinating and sometimes even planning what each will do as they act together toward a common end, attending to things jointly as they do. Children are thus engaging not just in cultural learning, which depends on understanding others as intentional agents, but rather, by formulating joint goals and intentions, they are engaging in full-blooded cultural creation. Perhaps of special note in this regard, 1- to 2-year-old children also begin participating in collaborative pretense activities in which they and the adult create together a shared fictional reality based on their joint intentions and attention (Rakoczy et al., 2005).

The cognitive representations underlying truly collaborative activities must contain at least two hierarchical levels: a higher one for the shared goal and a lower one for the joint intentions – with at least two sets of action plans (roles) in the joint intentions. This means that the cartoons of Figure 2 are meant to be taken seriously. Human cognitive representations may include people and their intentional actions in the world, including joint intentions between self and other. As these are, in essence, representations of social engagements, we may call them "dialogic cognitive representations" (Fernyhough 1996). Dialogic cognitive representations are necessary not only for supporting certain forms of collaborative interactions on-line, but they are also necessary for the creation and use of certain kinds of cultural artifacts, most importantly linguistic and other kinds of symbols, which are socially constituted and bidirectional. Dialogic cognitive representations may be ontogenetically emergent in the sense that the individual interacts in certain ways with other intentional agents, and then internalizes these interactions (see sect. 5.2).

Importantly, dialogic cognitive representations pave the way for later cognitive achievements that may be called, very generally, "collective intentionality" (Searle 1995). That is, the essentially social nature of dialogic cognitive representations enables children, later in the preschool period, to construct the generalized social norms (e.g., truth) that make possible the conceptualization of individual beliefs and, moreover, to share those beliefs. Sharing beliefs is responsible for the creation of social-institutional facts such as money, marriage, and government, whose reality is grounded totally in the collective practices and beliefs of a social group conceived generally (Tomasello & Rakoczy 2003). Importantly, when children internalize generalized collective conventions and norms and use them to regulate their own behavior, this provides for a new kind of social rationality (morality) involving what Searle (1995) calls "desire-independent reasons for action."

4. Apes and children with autism

An interesting question in all of this is the manner in which our nearest primate relatives are able to understand and share intentions. Obviously, an answer to this question would help to shed light on the phylogeny of social cognition in the human species, but it also would help to shed light on its ontogeny as well – by providing a kind of general primate starting point that might serve to isolate the evolutionarily unique features of human social cognition. Children with autism, who do not understand or interact with other persons in the species-typical manner for biological reasons, provide another perspective on the process from the point of view of atypical development, which can also quite often help us to carve nature at its joints.

4.1. Great apes

4.1.1. Understanding intentional action. Nonhuman primates are clearly able to use a variety of cues to predict the behavior of others in familiar situations, and even to try to influence their behavior communicatively, which suggest that they understand conspecifics as animate agents who produce their behavior spontaneously (Tomasello & Call 1997). Experimentally, using the Woodward habituation paradigm (as reviewed in Section 2.1), Santos and Hauser (1999) found that some monkeys expect that people continue to reach for an object that they have previously gazed at – just like human infants.

With regard to the understanding of goal-directed action, there is currently a good bit of controversy. Povinelli and Vonk (2003) consider the understanding of goals and perceptions to be an instance of understanding mental states, and their view is that apes understand only behavior not mental states. In contrast, Tomasello et al. (2003) (revising the view expressed in Tomasello & Call 1997) argue that there are now new data which compel us to attribute to great apes the ability to understand intentional action in terms of goals and perceptions.

Of most importance, it seems that apes understand both trying and accidents, in which the desired result never happens (see sect. 2.2). With regard to trying, Call et al. (2004) tested chimpanzees in a food-giving context similar to that of the Behne et al. study with human infants (described in Section 2.2). Specifically, a human began giving food to an ape through a hole in a Plexiglas wall, but then sometimes brought out a piece of food and either refused to give it to the ape (unwilling) or else attempted to give it to the ape unsuccessfully (unable). Similar to human 9- and 12-month-olds, chimpanzees gestured more and left the area earlier when the human was unwilling than when he was unable – in which case they tended to wait patiently throughout his well-meaning but unsuccessful attempts. The chimpanzees apparently understood the behavior of the human in the unable conditions as persistent attempts (trying) to give them food.[4]

With regard to accidents, comparisons of one pair of conditions in the Call et al. (2004) study also suggested that apes understand when someone is trying to give them something but clumsily failing. That is, apes also waited patiently when the human was making a good-faith, but clumsy and unsuccessful, effort. In addition, Call and Tomasello (1998) tested apes' ability to distinguish purposeful from accidental actions in a different paradigm. They trained subjects to associate a marker situated on top of one of three opaque buckets with the location of hidden food. In test trials, a human then placed the marker on one of the buckets purposefully, but either before or after this he let the marker fall accidentally onto one of the other buckets. Apes as a group chose the bucket that had been marked purposefully.

Chimpanzees also understand that others see things. They follow conspecific gaze direction to external targets (Okamoto et al. 2002; Tomasello et al. 1998), they check back with the looker (and eventually quit looking) if nothing is there (Call et al. 1998; Povinelli & Eddy 1996; Tomasello et al. 2001), and they even follow the gaze direction of humans to targets behind barriers (Tomasello et al. 1999). Chimpanzees also know that what others see affects what they do. Thus, Hare et al. (2000, 2001) placed a dominant and a subordinate individual into competition with each other over food – with some pieces of food visible to both individuals and some visible only to the subordinate chimpanzee. By pursuing most often the piece of food hidden from the dominant's view, subordinates demonstrated that they knew what the dominant could and could not see. And, importantly, the subordinates knew what this meant for the dominant's goal-directed action: if the dominant could see the food or had seen it just before, subordinates could infer that she would go for it (whereas they would not make this inference if what she saw was instead a rock). It is noteworthy that a monkey species did not behave like chimpanzees in the Hare et al. (2000) paradigm, and so perhaps this understanding is confined to apes (Hare et al. 2003).

It would thus seem that, at least on one reasonable reading of the data, some great apes understand at least some aspects of intentional action and perception. Apes understand that others have goals and behave toward them persistently, and that this is governed by what they perceive. This is still not an understanding of the more mental dimensions of intentional action, however – specifically those that have to do with the decision-making process by which the actor generates action plans and, based on a rational assessment of reality, chooses one to enact in intentional action. There is so far no evidence that apes understand this more mental dimension of the process, but at the moment there are no good tests of this – especially since imitation is not a very good way of investigating apes' social cognition. Indeed, many studies of imitation have shown that in response to a demonstration, apes tend to reproduce the result in the environment (*emulation learning*) and pay very little attention to the actual intentional actions of the demonstrator (see Tomasello 1996 for a review). This failure to engage in humanlike processes of cultural learning may be considered further evidence that apes are not so attuned to action plans or intentions.

4.1.2. Shared intentionality. Despite this sophistication in understanding many important aspects of intentional action, apes still seem to lack the motivations and skills for even the most basic forms of sharing psychological states with others. Thus, while ape infants interact with their mothers dyadically and are responsive to them behaviorally (Maestripieri & Call 1994) and they may even show some maternal gazing and social smiling (Mizuno & Takashita 2002; Tomonaga et al. 2004), there are no observations of anything like protoconversations between adults and infants. Personal observations of the authors suggest that although all primates display similar social emotions in terms of attachment between babies and mothers, human infants and mothers possess a much larger behavioral repertoire for expressing a much wider range of emotions in their social interactions than do other apes (e.g., laughing, crying, cooing, smiling) – especially expressions of positive emotions serving to enrich the dyadic emotional engagement between mother and child.

Similarly, apes engage in very few triadic interactions with others around objects. They beg food from one another, and youngsters' play sometimes incorporates objects. But systematic observations of chimpanzee and bonobo mothers and infants with objects reveals very little triadic engagement, and none that appears to involve a shared goal (Bard & Vauclair 1984; Tomonaga et al. 2004). When apes interact with humans, they engage in more triadic interactions, but these interactions are still discernibly different from those of human mothers and babies. For example, Carpenter et al. (1995) observed human 18-month-olds as well as chimpanzees and bonobos in interaction with an adult human and some objects. In this situation, all three species interacted with objects and simultaneously monitored the adult human's behavior reasonably frequently. However, there were also important differences. Human infants spent far more time in joint attentional episodes, and their looks to the face of the adult were, on average, almost twice as long as those of the apes. Infant looks were also sometimes accompanied by smiles, whereas apes do not smile. These differences gave the impression that the ape's look to the adult was a checking look (to see what the adult was doing or was likely to do next), whereas the infant's look to the adult was a sharing look (to share interest). One interpretation of this pattern of observations is that although apes know that others have goals and perceptions, they have little desire to share them. They can interact with others triadically around objects, but they do not engage with others in shared endeavors with shared goals and experiences.

With regard to collaborative engagement, chimpanzees join one another in agonistic interactions within the group (so-called coalitions and alliances), and they act together to defend the group from predators and other chimpanzee groups. But in these interactions each individual does basically the same thing, they just do it in concert without any discernible coordinated plans. The most complex cooperative activity of chimpanzees is group hunting, in which two or more males seem to play different roles in corralling a monkey (Boesch & Boesch 1989). But in a reinterpretation of participant behavior over time in these hunts, some observers have characterized this activity as essentially identical to the group hunting of other social mammals such as lions and wolves (Cheney & Seyfarth 1990; Tomasello & Call 1997). Although it is a complex social activity, as it develops over time each individual simply assesses the state of the chase at each moment and decides what is best for it to do. There is nothing that would be called collaboration in the narrow sense of joint intentions and attention based on coordinated plans. In experimental studies (e.g., see Crawford 1937; Chalmeau 1994), the most complex behavior that can be extracted is something like two chimpanzees pulling a heavy object in parallel, and during this activity almost no communication among partners is observed (Povinelli & O'Neill 2000). There are no published experimental studies – and several unpublished negative results (two of them ours) – in which chimpanzees collaborate by playing different and complementary roles in an activity.

In general, it is almost unimaginable that two chimpanzees might spontaneously do something as simple as carry something together or help each other make a tool, that is, do something with a commitment to do it together and to help each other with their role if needed. Indeed, in a recent study, Hare and Tomasello (2004) found that in a single food-finding task structured as either competition or cooperation, chimpanzees performed much more skillfully in the competitive version. Nor does ape communication seem to be collaborative in the same way as human communication. Most basically, there is very little communication about third entities (topics), and there are no signals serving a declarative or informative motive. Apes do not point, show, or even actively offer things to conspecifics.[5] Also, Tomasello (1998) argues and presents evidence that chimpanzee signals are not really bidirectional in the sense that sender and receiver both know that either could play either role (i.e., they do not know it is the same signal when they send it as when they receive it).[6] There are also a number of experimental studies demonstrating that apes are not able to understand communicative intentions as manifest in such acts as pointing or placing a marker to indicate the location of food (for a review, see Call & Tomasello, 2005). Finally, in no case does there seem to be any kind of negotiating over intended meaning, requests for clarification, or other kinds of negotiation (Liebal et al., 2004). In general, although chimpanzee groups in the wild do have different behavioral "traditions" (Boesch 1996), chimpanzees' relatively modest skills of collaboration would not seem to be of the type necessary for cultural creation of the human kind.

The overall conclusion would thus seem to be that although apes interact with one another in myriad complex ways, they are not motivated in the same way as humans to share emotions, experiences, and activities with others of their own kind. They do not look to others and smile in order to share experience triadically, they do not invite others to share interest and attention via declarative gestures, they do not inform others of things or help them in their efforts, and they do not engage with others in collaborative activities with shared goals and joint intentions. But what if they are raised in a human cultural environment in which they are encouraged to engage in collaborative activities and communicate with symbols? The basic answer is that apes raised in such environments become more like humans than their wild conspecifics, but they do not turn into humans (Call & Tomasello 1996). Thus, Savage-Rumbaugh (1990) reports that the bonobo Kanzi participates regularly in social activities such as preparing food and playing with toys. But it is not clear whether he has the kind of commitment to these activities as joint endeavors that characterizes human collaboration, and there is no evidence that he understands the role of the other or supports him in it. In his mainly imperative attempts at communication, Kanzi does not simply share interest with or inform others, and he does not negotiate over meaning or support the other collaboratively in the communication process through requests for clarification or adjustments for listener knowledge (Greenfield & Savage-Rumbaugh 1991).

4.2. Children with autism

4.2.1. Understanding intentional action. Children with autism clearly understand other persons as animate beings who produce behavior spontaneously, as evidenced by their social behavior in general. In addition, in the few nonverbal studies that have been conducted, these children show some signs of understanding that others have goals and that others see things. Thus, 3- to 4-year-old children with autism look more to an adult's face following ambiguous actions than unambiguous actions – presumably in an attempt to discern the adult's goal (Carpenter et al. 2002; Charman et al. 1997; though see Phillips et al. 1992 for a negative finding). Using imitation tasks, Carpenter et al. (2002) found that 3- to 4-year-old children with autism not only imitated an adult's unusual action, such as turning on a light with the head, but also looked to the light in anticipation, seemingly indicating their appreciation of the goal-directed nature of this action. With regard to trying, two studies using versions of Meltzoff's (1995) behavioral reenactment procedure found no clear impairment for children with autism (Aldridge et al. 2000; Carpenter et al. 2001), suggesting their appreciation of the persistent nature of goal-directed action. Findings are mixed with regard to the cultural learning skills of children with autism (e.g., see Rogers 1999 for a review). But Hobson and Lee (1999) found that children with autism imitated the particular style of a demonstrator's actions less often than did other children. This might suggest, among other things, that they are less skilled at analyzing intentional action down the hierarchy of means.

In terms of an understanding of perception, children with autism show clear deficits in tests of spontaneous gaze following, but, interestingly, when instructed to do so, they can report what the other person is looking at (Leekam et al. 1997). There are, to our knowledge, no direct tests of their ability to follow gaze around barriers or in any other way to demonstrate whether they understand that others do not just look at things but actually see or attend to things. One plausible hypothesis, then, is that at least some children with autism (perhaps on the high-functioning end of the spectrum) understand that others have goals and behave toward them persistently (and that others see things). However, they still may not understand the decision-making process by which an actor rationally chooses among potential behavioral means to generate intentional action; this has not been explicitly tested as yet.

4.2.2. Shared intentionality. Unfortunately, the skills children with autism have for understanding intentional action and perception do not translate into motivations and skills for sharing psychological states with others – with even the name of the disorder suggesting this deficit. With regard to shared dyadic engagement, Hobson (2002) reviews much evidence that children with autism have special problems in recognizing, understanding, and sharing emotions with others, and so they do not seem to engage in protoconversations.

Deficits with regard to shared triadic engagement and joint attention are so pervasive in children with autism that they actually represent diagnostic criteria. Perhaps of special importance, children with autism show very little coordinated joint engagement, and initiate very few bids for joint attention with others by declaratively pointing to or showing objects (e.g., see Baron-Cohen 1989; Charman et al. 1997; Mundy & Willoughby 1996) – which indicates most clearly their lack of motivation. They also rarely respond to others' bids for joint attention (e.g., see Leekam et al. 1997). With regard to collaborative engagement, children with autism engage in relatively little cooperative play with peers and in general collaborate with others very little (Lord 1984), and there is little evidence of role reversal or helping others in their role (Carpenter et al. 2005). Linguistic communication and the use of symbols is another problem area for children with autism, and their impaired ability to signal noncomprehension and make appropriate repairs to their own linguistic messages to help others are well documented – suggesting that their communication is not fully collaborative (Loveland et al. 1990). Hobson (2002) argues and presents evidence that in fact all of these problems may be traced back to problems with emotional relatedness, that is, a deficit in the normal human motivation to share emotions, experiences, and actions with other persons. The outcome is that, although there may be a few unusual individuals, the vast majority of children with autism do not participate in the cultural and symbolic activities around them in anything like the normal way.

4.3. Summary

Great apes and children with autism are clearly not blind to all aspects of intentional action. Contrary to some previous accounts, both apes and some children with autism do appear to understand actions as goal directed if not fully intentional; that is, they understand that others have goals, persist toward them, and perceptually monitor the process. This means that both of them show some skills of social learning, though not as powerful or pervasive as those of human 1- and 2-year-olds. However, neither apes nor children with autism follow the typical human developmental pathway of social engagement with other persons. Neither of them engages with other persons in shared dyadic engagements (protoconversations), shared triadic engagements (joint actions), or collaborative engagements (with joint intentions and attention). And there does not seem to be anything like a declarative motivation simply to share attention with others or to inform others or to help others, anywhere in sight. In general, it seems that neither apes nor children with autism have – at least not to the same extent as typically developing human children – the motivation or capacity to share things psychologically with others. This means that they both have very limited skills for creating things culturally with other persons.

5. Two hypotheses

Based on all of these data, our proposal is that in addition to understanding others as intentional, rational agents, humans also possess some kind of more specifically social capacity that gives them the motivation and cognitive skills to feel, experience, and act together with others – what we may call, focusing on its ontogenetic endpoint, shared (or "we") intentionality. As the key social-cognitive skill for cultural creation and cognition, shared intentionality is of special importance in explaining the uniquely powerful cognitive skills of *Homo sapiens*. And so our question now is Where does this capacity for shared intentionality come from phylogenetically and ontogenetically?

5.1. A phylogenetic hypothesis

Primates are intensely competitive creatures. By most accounts, the social-cognitive skills that distinguish primates from other mammals evolved mainly in the context of competitive social interactions, and so, following Humphrey (1976), primate social cognition has been characterized by appellations such as primate politics (de Waal 1982) and Machiavellian intelligence (Byrne & Whiten 1988). In experimental comparisons, at least some primate species show their most sophisticated social-cognitive skills in competitive rather than in cooperative situations (Hare and Tomasello 2004; Hare et al. 2000; 2001).

Our proposal is that, in addition to competing with others (and coordinating with others generally, like all social animals), humans evolved skills and motivations for collaborating with one another in activities involving shared goals and joint intentions/attention. At some point – perhaps heralding the emergence of modern humans some 150,000 years ago – individuals who could collaborate together more effectively in various social activities came to have a selective advantage. This may have happened within groups, in a manner analogous to the hypothesis of Wrangham (1980), who argues that because many primates forage for patchy resources such as fruit, and patchy resources may be easily dominated by a small group of individuals to the exclusion of others, some primates have evolved social systems in which small groups (e.g., matrilineal kin groups as well as more temporary coalitions and alliances) act together so as to compete with groupmates for valued resources (see also van Schaik 1989). Humans may simply have pushed this process – small bands acting together to compete with other bands in their group – a bit further by turning "acting together" into collaborating. But the evolution of humans' unique skills of collaboration may also have happened between groups. Thus, it is also possible that some kind of group-level selection played an important role in the evolution of these collaborative activities, as some change in the ecology of *Homo* made it more likely that entire groups with many collaborators outcompeted other groups with fewer collaborators (Sober & Wilson 1998).

The key cognitive substrate required for skillful collaboration is the ability to read intentions. Although intention reading may be helpful in competitive interactions, it is not absolutely necessary – since in competition I care mainly about what you *do*. That is to say, in competitive interactions, the interactants do not have goals about others' intentional states; the situation is that we both have the "same" goal (e.g., we both want that piece of food), and the key thing is that I anticipate what you will do next. In contrast, collaborative interactions require interactants to have goals about others' intentional states so that the requisite shared goals and plans may be formulated. Thus, in collaborative interactions, we are faced with the so-called coordination problem from the outset: to even get started, we must somehow coordinate or negotiate so that we end up with a shared goal (which we did not have to start with; see Levinson 2000). Then, in addition, to collaborate effectively, we must mesh our action plans at least some of the way down the hierarchy – and this requires some communication about those plans, at least to some degree ahead of time.

Phylogenetically, it is possible that the selection process favoring collaborative individuals worked on variation in intention reading of the type currently represented in the great apes. But, more likely, earlier members of the genus *Homo* developed especially complex skills of intention reading in the context of the imitative learning of complex tool-using and tool-making activities – which require a hierarchical analysis of goals and plans – so that the selection process on modern humans was working with individuals already especially adept at discerning the intentional structure of action. This account would also explain why it is that modern humans seem to be so much more skillful at imitation than other apes – especially when the task requires a means-ends analysis of the observed behavior (Tomasello 1996).

The key motivational substrate required for collaboration is the motivation to share feelings, experiences, and activities with other persons – where again sharing means having psychological states that include within them as content the psychological states of others. Perhaps following Hare and Wrangham (2002), we might propose a first step of increased within-group tolerance, as humans (and to some degree bonobos) essentially "domesticated" themselves relative to the *Pan-Homo* common ancestor of 6 million years ago – ostracizing overaggressive and less-tolerant groupmates. But this is not enough. In addition, collaborative activities require more active motivations for sharing emotions, experience, and intentional actions with others. For example, communicating only to share interest in things and communicating only to share information seem to be uniquely human activities (what Dunbar 1996 calls gossiping), and imitation for purely social motivations – not just to accomplish goals but to be like others – is a key component in the transmission of human culture (Tomasello 1999b). In addition, it may even be that humans have some "altruistic" motives for helping others in the sense that they are motivationally built for strong reciprocity, in which their behavior is governed by social norms of "fairness" (Boyd et al. 2002; Gintis et al. 2003).[7] Again, it is possible that the selection process favoring collaborative individuals worked on variation in the motivation to share of the type currently represented in other apes. But it is also possible that by the time this selection process took place, *Homo* had already evolved some new social motivations, perhaps in the context of nuclear families (Wrangham et al. 1999).

We thus envision that the individuals of some premodern human population, possessing something like modern-day chimpanzee "culture" (Boesch 1996), evolved the skills and motivations of shared intentionality, which enabled especially complex forms of collaboration and resulted eventually in modern human cultural organization. It is possible that individual selection could do the whole job, as in many cases collaborative actions have mutualistic benefits to both participants. Or there may also have been, in addition, some form of group-level selection (Sober & Wilson 1998) or cultural group selection (Boyd et al. 2002), relying on social norms of strong reciprocity and cultural conformity. The coevolution of skills of intention reading and collaboration then enabled – via cultural-historical processes involving the ratchet effect – creation of the many collective artifacts and social practices that constitute particular human cultures and that structure the cognitive ontogenies of developing youngsters. Our proposal thus supplements a Machiavellian account of human cognitive evolution, which emphasizes only competition, with a Cultural account that emphasizes in addition the importance of collaboration, cultural-historical processes, and strong reciprocity based on social norms.

5.2. An ontogenetic hypothesis

If our phylogenetic hypothesis is correct, selection for good collaborators means selection for individuals who are (1) good at intention reading and (2) have a strong motivation to share psychological states with others. Our ontogenetic hypothesis is that it is precisely these two developing capacities that interact during the first year of life to create the normal human developmental pathway leading to participation in collaborative cultural practices.

As for the first, intention-reading, line of development, there have been a number of proposals to the effect that this skill is a hardwired and modular part of the human perceptual system. Just as humans automatically see certain perceptual sequences as causal (Leslie 1984; Michotte 1963), they automatically see certain actions performed by animate agents as goal directed. Gergely and Csibra (2003) have proposed that human infants possess an action-interpretation system that perceives humanlike action as teleologically directed to a goal from the second half of the first year of life; independently developing is a reference-interpretation system concerned with following gaze and the like (Csibra 2003). Baron-Cohen (1995) proposes something similar, with two early developing innate modules involving the perceiving of goals and eye gaze direction. Soon after the first birthday, a "shared attention mechanism" emerges, taking the two earlier modules as inputs.

Although our view shares some features with these views, there are two important differences. First, we do not see infants' understanding of goals/intentions and perception/attention as blocked off from each other in a modular fashion. Indeed, much of the evidence we have presented here suggests that in attempting to understand what others are doing and why they are doing it, infants comprehend intentional action and perception as an integrated system (i.e., as a kind of control system). They display such an integrated understanding from 9 months of age when they know that an actor pursues goals persistently (until he perceives that the world matches his goal) and also engage with other persons triadically around external objects – where they must infer people's perceptions from their goals and their goals from their perceptions. In general, we do not see how an observer can understand goal-directed action (much less rational action) without understanding a perceiving organism who monitors the world for signs of success, failure, obstacles, and so forth.

Second, we believe that to understand the origins of a human cognitive skill we must go beyond simply labeling it as "innate." Indeed, although we concur that understanding actions as goal directed is a biological adaptation, this says nothing about the ontogenetic process. It is very unlikely, in our view, that a human or ape kept in social isolation for the first year of life would suddenly understand others as goal-directed or intentional agents on its initial encounter with them; presumably, the developmental pathway for understanding intentional action depends on species-typical social interactions early in ontogeny. This does not necessarily mean, however, any specific experiences. Thus, Kaye (1982) proposes that to understand intentions infants must themselves be treated by adults as intentional, in the sense that adults interpret their actions in adultlike terms and provide various types of feedback to this effect. The problem with this more specific hypothesis is that there seems to be fairly wide cultural variation in how infants are treated by adults – with adults in some cultures not really treating infants as fully intentional – and, by all accounts, all children in all cultures develop an understanding of others as intentional agents.

As for the second, sharing, line of development, theorists such as Trevarthen (1979), Bråten (2000), and especially Hobson (2002) have elaborated the interpersonal and emotional dimensions of early human ontogeny in much more detail than we have here. We mostly agree with their accounts, but we find that they do not give sufficient attention to the other, intention-reading, line of social-cognitive development. Our proposal is that the uniquely human aspects of social cognition emerge only as uniquely human social motivations interact with an emerging, primate-general understanding of animate and goal-directed action – which then transforms the general ape line of understanding intentional action into the modern human line of shared intentionality.

Although the precise nature of this interaction is not entirely clear, our general view is that infants begin to understand particular kinds of intentional and mental states in others only after they have experienced them first in their own activity and then used their own experience to simulate that of others (Tomasello 1999b; for experimental evidence supporting this view, see Sommerville & Woodward 2005). However, contrary to our previous view, we do not think that simple "identification with others" is a sufficient basis for the simulation process – certainly not if we mean bodily identification, as there is now evidence that neonatal chimpanzees engage in the same kind of facial mimicking as human infants (Myowa 1996; Myowa-Yamakoshi et al. 2004), and even some species of birds are good at copying actions (e.g., see Zentall 1996). And so we would speculate at this point that more deeply psychological levels of identification with others – of a kind sufficient to enable individuals to simulate the intentional and mental states of others on analogy with their own – depend crucially on the skills and motivations for interpersonal and emotional dyadic sharing characteristic of human infants and their caregivers (Hobson 2002).

Again, one can imagine that a species-typical social environment, involving human-typical social interactions with other persons, is required for the emergence of the sharing motivation and its related skills of social engagement. But, again, some theorists have proposed that some kinds of specific experiences are necessary. For instance, Stern (1985) proposes that parents must "mirror" back to infants their own emotions or behaviors, and Gergely (2001) posits an especially important role for certain kinds of social contingencies in terms of timing. But, again, it is not clear that children in all cultures receive such experiences, or that children who are deprived of them end up unable to share psychological states with others. And so the ontogenetic process for sharing emotions and intentions with others may be fairly robust in the face of different particular human social environments.[8]

Based on this analysis and on our review of the developmental research in Sections 2 and 3, then, our proposal for the early developmental pathway characteristic of human social cognition is thus:

- Young infants understand other persons as animate agents and so share emotions and engage with them dyadically;
- 9-month-olds understand other persons as goal-directed agents and so share goals (and perception) and engage with them triadically; and
- 14-month-olds understand other persons as intentional agents and so share intentions (and attention) and engage with them collaboratively (so creating, via internalization, dialogic cognitive representations).

This pathway is a synergistic product of the general ape line of understanding intentional action, unfolding from 0 to 14 months, and the modern human motivation to share psychological states with others, present from very early in human ontogeny. Figure 3 provides a schematic overview of this account. As noted above in this section, there has been almost no research – not even training studies or correlational studies – that establishes a solid relationship between any kind of particular social experience infants might have and individual differences in the unfolding of this developmental pathway. In the absence of such studies, we might tentatively conclude that this is a very robust, heavily canalized ontogenetic pathway in humans that emerges in all "normal" human environments.

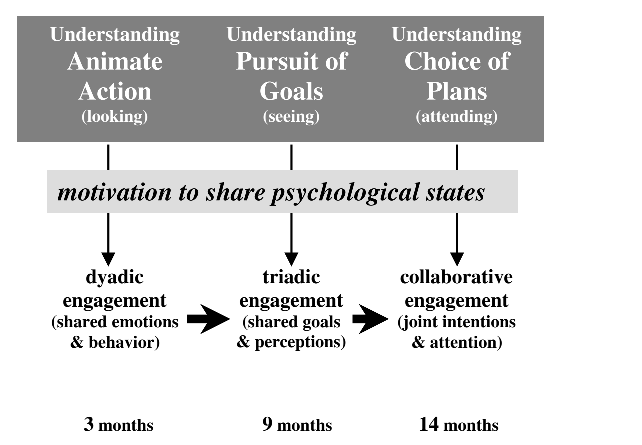

Figure 3. Ontogenetic pathway for human social engagement as a joint product of the understanding of intentional action and the motivation to share things with others psychologically.

What results from this developmental process, early in the second year of life, is a new form of cognitive representation, what we have called dialogic cognitive representations, and they enable children's participation in truly collaborative cultural practices such as linguistic communication and other forms of symbolic interaction. Dialogic cognitive representations include and go beyond theoretical constructs such as "identification with others" (Hobson 1993; Tomasello 1999b), the "like me" stance (Meltzoff & Gopnik 1993), and "self-other equivalence" (Barresi & Moore 1996) – which may be ontogenetic forerunners. That is to say, they capture the fact that the child both knows that she is in some sense equivalent to others – actors can substitute for one another in acts of imitation and role reversal – but at the same time she is different from others. Dialogic cognitive representations thus have built into them the functional equivalence (though not identity) of different participants in activities, one of whom may be the self, but they have additional aspects (e.g., intentions about the other's intentions) deriving from the motivation to share psychological states with others.

At this point, we are in no position to offer a specific hypothesis about how dialogic cognitive representations are created ontogenetically beyond the general claim that the sharing of psychological states engaged in by human infants and caregivers is in some way internalized in Vygotskian fashion. Perhaps a bit more specifically, we might hypothesize that in understanding an adult's intentional actions, including those directed toward her, at the same time that she experiences her own psychological states toward the other, the child comes to conceptualize the interaction simultaneously from both a first and third person perspective (see Barresi & Moore 1996) – forming a "bird's-eye view" of the collaboration in which everything is comprehended in a single representational format.[9] During months and even years of such interactions, from ages 1 to 5 and beyond, children come to construct in dialogic fashion such things as social norms and their constitutive conventional practices and individual beliefs. This enables them to participate in and contribute to the collective social practices and institutions around them, that is, to participate in and contribute to the collective intentionality of a human culture.

6. Conclusion

Human cognition sticks out like an elephant's trunk, a giraffe's neck, a peacock's tail. It is one form of primate cognition, but it seems totally unique as people go around talking and writing and playing symphonies and doing math and building buildings and engaging in rituals and paying bills and surfing the web and creating governments and on and on. Also unique in the animal kingdom, human cognition is highly variable across populations, as some cultures have complex foraging and navigational techniques whereas others have very few of these, and some do algebra and calculus whereas others have very little need for complex mathematics. And so the biological adaptation we are looking for is one that is rooted in primate cognition but then provides humans with the cognitive tools and motivations to create artifacts and practices collectively with members of their social group – that then structure their and their offspring's cognitive interactions with the world. We are thus looking for a small difference that, by creating the possibility of culture and cultural evolution, made a big difference in human cognition.

Our proposal for this "small difference that made a big difference" is an adaptation for participating in collaborative activities involving shared intentionality – which requires selection during human evolution for powerful skills of intention reading as well as for a motivation to share psychological states with others. In ontogeny, these two components – the understanding of intentional action and the motivation to share psychological states with others – intermingle from the beginning to produce a unique developmental pathway for human cultural cognition, involving unique forms of social engagement, symbolic communication, and cognitive representation. Dialogic cognitive representations, as we have called them, enable older children to participate fully in the social-institutional-collective reality that is human cognition.

There are two other main theoretical contenders for what makes human cognition unique in the animal kingdom. First, of course, many theorists point to language, and without a doubt language must play a central role in all discussions of the evolution of human cognition. But saying that only humans have language is like saying that only humans build skyscrapers, when the fact is that only humans (among primates) build freestanding shelters at all. Language is not basic; it is derived. It rests on the same underlying cognitive and social skills that lead infants to point to things and show things to other people declaratively and informatively, in a way that other primates do not do, and that lead them to engage in collaborative and joint attentional activities with others of a kind that are also unique among primates. The general question is What is language if not a set of coordination devices for directing the attention of others? What could it mean to say that language is responsible for understanding and sharing intentions, when in fact the idea of linguistic communication without these underlying skills is incoherent. And so, while it is true that language represents a major difference between humans and other primates, we believe that it actually derives from the uniquely human abilities to read and share intentions with other people – which also underwrite other uniquely human skills that emerge along with language such as declarative gestures, collaboration, pretense, and imitative learning (Tomasello 2003). Of course, later in ontogeny, there may be some cognitive achievements possible only with the support of the linguistic version of dialogic cognitive representations, which embody in special ways the different perspectives and construals that people may take on things (Lohmann et al., 2005).

The other major contender for what makes human cognition unique is theory of mind. Our proposal is of course one variant of this, and indeed we would argue that the full understanding of intentional action, including its rational and normative dimensions, involves some understanding of things mental. But when most people use the term theory of mind they mean the belief-desire psychology with which school-age children and adults operate. But this form of theory of mind is clearly derivative of more basic social-cognitive skills. Thus, Tomasello and Rakoczy (2003) argue and present evidence that while the understanding and sharing of intentions emerges ontogenetically in all cultural settings at around 1 year of age – with no known individual differences due to environmental factors – the understanding of beliefs emerges some years later at somewhat different ages in different cultural settings, and there is very good evidence that participating in linguistic communication with other persons (especially some forms of perspective-shifting discourse) is a crucial, perhaps even necessary, condition for its normal development. And so again, while the understanding of beliefs and desires is clearly a critical component in uniquely human cognition and culture, we do not believe it is basic, but rather it, too, is derived from the understanding and sharing of intentions.

Having argued that an adaptation for shared intentionality is more basic than other theoretical contenders such as language and theory of mind, we must also acknowledge that there could be other hypotheses about the origins of uniquely human cognition that are more basic still. For example, one could hypothesize that humans simply evolved larger brains with more computing power than other primates – maybe specifically a larger working memory that enables them to hold more things in mind simultaneously (e.g., see Olson & Kawamar 1999) – and that this was sufficient to create all the differences we see today between humans and other primates. Also, one could hypothesize a very simple difference in sociality between humans and other animals, such as the tendency to be responsive to the rewards, punishments, and direction of others in the social group (e.g., see Wilson 1999 on consilience). But in these cases we would argue that such nonspecific adaptations are not sufficient to get the job done. To get from primate social groups to human cultures and the collective cognition they embody, something like an adaptation for participation in collaborative activity is required – leading to selection for motivations and skills of shared intentionality and the cultural-historical processes these engender.

There is of course still much we do not know about all of this. We do not know with much precision the degree to which humans and other apes differ in their understanding of how others choose plans – the rational aspects of intentional action – since most of the studies done with infants cannot be so easily done with apes. We have very little specific knowledge about humans' motivation to share things psychologically with others, in this case because the most telling experiments (e.g., isolation experiments) would be unethical. We do not know exactly how much of an understanding of intentional action is necessary for children to participate in collaborative activities. And conversely, we do not know whether the kinds of collaborative activities that exist in cultures before children are born are a necessary or only a facilitative component in the ontogenetic process – or whether they play no effective role at all at the outset (though clearly they play a crucial role later). Our view is that to make progress on these and related questions we must focus our research efforts both on the individual cognitive skills required to understand intentional action, in all its many dimensions, and also, in equal measure, on the social motivations and dialogic representations that underlie the collaborative activities and collective artifacts that structure human culture and cognition.

NOTES

1. Note that in a situation in which current reality matches the desired goal state, the organism will not behave (it will have no desire to behave because its goal is already met). It is also possible that, in some cases, inaction is a good strategy for bringing the world in line with one's goals. That is, in some cases, inaction may be intentional action, an insight possible only if one considers all of the components of a control system working together.

2. This study – or any other with its same logic – has yet to be done with younger children. It might be argued that the study by Gergely et al. (1995) showed that infants know that organisms adjust their behavior to reality constraints in the form of obstacles. But the dishabituation methodology does not enable such an inference because the child does not have to choose an action plan herself (as in imitation studies). Thus, in that study, infants only needed to discriminate normal from abnormal behavior: goal-directed agents do not normally take circuitous routes to goals. (A similar argument applies to the study by Woodward & Sommerville 2000.)

3. In some accounts of shared intentionality, it is enough that we both have the same goal and know that we do (i.e., have mutual knowledge of the fact that we both have the same goal). But this is not enough; we might each want the box open and know that the other does also, but still not form a shared goal (perhaps we will compete to see who can open it). Further, it is also not enough simply to have goals about our behaving together. If I suggest we go to the movie together, my desire is not that you come because your mother forced you to but because you want to – I want us to have a shared commitment. (Note, however, that because of the hierarchical structure of action, there may exist many mixed cases in which you collaborate reluctantly because of competing goals and so forth.)

4. In a different experimental paradigm, Myowa-Yamakoshi and Matsuzawa (2000) and Call et al. (2005) both used Meltzoff's (1995) behavioral reenactment procedure (involving trying and failed attempts) with chimpanzees. Both found that chimpanzees, like children, performed the target action equally as often when they saw a failed attempt as when they saw the completed action. However, in both studies, chimpanzees also performed the target action at high levels in a baseline condition containing no demonstration at all, which seriously limits what can be concluded about the subjects' understanding of the modeled action.

5. Vervet monkey alarm calls and the like do not need to be interpreted as referential, and indeed individuals have very little control over their production at all (Owren & Rendell 2001). Moreover, there is no evidence that any ape species uses such calls (Tomasello & Call 1997).

6. Although sometimes presented in this way, the study by Povinelli et al. (1992) has other interpretations not involving role reversal (Tomasello & Call, 1997).

7. Evidence for this view is provided by experimental studies in paradigms such as (1) the ultimatum game in which individuals offer more money to others than would be beneficial from a selfish viewpoint, at least partly because this seems like the "fair" thing to do (Gintis et al. 2003); and (2) experimental games in which individuals go to great lengths to punish others who are not being "fair" even when this punishing act could not possibly lead to future benefits for the punisher that outweigh the costs (on altruistic punishing, see Fehr & Gächter 2002).

8. That is, at least with respect to basics – specific environmental differences may of course create important individual differences, some considered atypical or even pathological.

9. Barresi and Moore (1996) are focused on a different problem, claiming that in order to attribute psychological states to others at all the infant must first interact with them in situations in which they both have similar psychological reactions. We are focused on collaboration and dialogic cognitive representations among agents who already understand one another intentionally, and our hypothesis is that the child internalizes these interactions into cognitive representations that encompass simultaneously both first-person and third-person perspectives.

Open Peer Commentary

Language first, then shared intentionality, then a beneficent spiral

Derek Bickerton

Department of Linguistics, University of Hawaii, Honolulu, HI 96822.
derbick@hawaii.rr.com www.derekbickerton.com

Abstract: Tomasello et al. give a good account of how shared intentionality develops in children, but a much weaker one of how it might have evolved. They are unduly hasty in dismissing the emergence of language as a triggering factor. An alternative account is suggested in which language provided the spark, but thereafter language and shared intentionality coevolved.

Nobody could dispute Tomasello et al.'s major premise – that shared intentionality forms the basic infrastructure of the uniquely human capacity to collaborate. Nor would one argue against their conclusion that apes possess at least the rudiments of intentionality, plus a spot of primitive mind reading (although parents of autistic children may not be too happy about having them compared with apes). The authors are right to look for a missing, uniquely human ingredient.

But in their search for such an ingredient they walk right over it at the very beginning. In their introduction, they state that collaborative acts are "structured by shared symbolic artifacts, such as linguistic symbols and social institutions, facilitating their 'transmission' across generations." The question that immediately arises is whether language merely structures these acts and facilitates their transmission, or whether it is itself the missing ingredient, both a necessary and sufficient condition for the capacities at issue. If so, looking for another cause would violate Occam's razor.

We are told that we are biologically adapted for shared intentionality, but the authors devote most of the target article to an ontogenetic rather than a phylogenetic analysis, and offer only a skeleton outline of the evolutionary processes involved. Some of the developmental studies they cite are indeed impressive; perhaps most of all that by Tomasello and Haberl (2003), which suggests we should take another look at accounts claiming children under age 4 have no theory of mind. However, unless one swallows Haeckel wholesale, there are no grounds for assuming that the ontogenetic order of the developments described precisely mirrors that of their phylogenetic emergence.

Indeed, when the authors attempt an evolutionary account, they are reduced to implausible just-so stories. It is suggested that early humans may have merely intensified a primate trait in which small bands competed for scarce resources, leading to greater intragroup cooperation, or that some "change in the ecology of *Homo*" caused groups that contained more collaborators to outcompete groups that contained fewer. What scarce resources? What change in ecology? Why did humans, rather than any of the other primate species, begin to collaborate? Why did their collaboration grow to such a vast extent while other primates still do not collaborate at all? No answers are provided. Moreover, there is not one shred of evidence in the entire paleontological record for the kind of scenario sketched here.

But some such scenario was more or less forced on the authors by their decision that language could not have been the missing ingredient. In section 6, they make the bald assertion that "Language is not basic; it is derived." They ask, "What is language if not a set of coordination devices for directing the attention of others?" Well, any number of things, but most importantly an entirely novel means of structuring experience and representing the world (Bickerton 1990).

It is claimed that the notion of linguistic communication without understanding and sharing intentions is incoherent. But what about protolinguistic communication (Bickerton 1990)? The authors would be right if the understanding and sharing of intentions, on the one hand, and language, on the other, had suddenly emerged ready-made; if we did not understand and share intentions the way we do now, language as we know it now would indeed be unworkable. But none of these things dropped from the skies in their current state. All evolved, presumably from very humble beginnings, and it is in dealing with these beginnings that the article is weakest.

The word *coevolution* is tossed around pretty freely these days, but here is where a really strong case could be made for it. Very little understanding or sharing of intentions – perhaps little if any beyond what contemporary apes possess – would have been required to comprehend and act on the kind of single-unit utterances with which language must have begun. (Or do the authors propose that our ancestors suddenly started spouting full grammatical sentences, like the infant Lord Macaulay?) But once the process began, every increment in linguistic skill could lead to an increase in shared intentionality, and vice versa.

The question is, of course, a chicken-and-egg one. Did language trigger shared intentionality, or vice versa? One interesting difference between the two lies in the fact that shared intentionality had primate precursors, whereas language didn't. Tomasello et al. themselves list some of those precursors in section 4.1.1; the differences between apes and children that they point out in section 4.1.2 are mainly matters of degree. Language, however, differed radically and qualitatively from anything that had gone before. It seems plausible to suppose that the radical difference triggered the spurt in the more scalar one rather than vice versa.

A commentary with a thousand-word cap hardly gives room to flesh out an alternative scenario. However, I would urge the authors to consider the kind of coevolutionary account I have merely hinted at here. When all is said and done, is it too trivially obvious to ask what force could have driven shared intentionality more effectively than the ability to tell one another our intentions?

Joint cooperative hunting among wild chimpanzees: Taking natural observations seriously

Christophe Boesch

Department of Primatology, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. **boesch@eva.mpg.de**

Abstract: Ignoring most published evidence on wild chimpanzees, Tomasello et al.'s claim that shared goals and intentions are uniquely human amounts to a faith statement. A brief survey of chimpanzee hunting tactics shows that group hunts are compatible with a shared goals and intentions hypothesis. The disdain of observational data in experimental psychology leads some to ignore the reality of animal cognitive achievements.

In the past, philosophers and scientists have regularly proposed new definitions of human uniqueness based on their personal convictions and intuitions of what animals are or are not able to do. Nowadays, over 45 years of field studies on wild chimpanzees provide a wealth of observational data against which to confront these preconceptions. In this sense, it is more than surprising to find only a single reference to animal field data in Tomasello et al.'s long citation list. Not surprisingly, their portrayal of cooperative hunting in chimpanzees reminds one of the old philosophers' claims. This is especially disappointing in that their proposition that the ability to share goals and intentions is a uniquely human capacity rests squarely on the assumption that no other species can do so.

I will briefly outline an analysis of the hunting behaviour among wild chimpanzees showing that individual hunters' behaviour is noticeably compatible with sharing goals and intentions. Hunting has been observed in all chimpanzee populations studied so far, and large differences in hunting strategies have been documented, especially in the propensity to hunt in collaborative groups (Boesch 1994a; 1994b; Mitani & Watts 1999; 2001; Nishida et al. 1992; Stanford 1998; Stanford et al. 1994a; 1994b; Watts & Mitani 2000; 2002). Natural observations can address only the question of performance, but we know from human observations that comprehension often exceeds performance (Birch & Bloom 2004; Keysar et al. 2003). During 77% of the 274 group hunts followed, Taï chimpanzees performed four complementary hunting roles (Fig. 1). Briefly: The *driver* initiates the hunt by slowly pushing the arboreal prey in a constant direction, *blockers* climb trees to prevent the prey from dispersing in different directions, the *chaser* may climb under the prey and by rapidly running after them try a capture, and the *ambusher* may silently climb in front of the escape movement of the prey to block their flight and close a trap around the prey (Boesch 1994a; 2002; Boesch & Boesch-Achermann 2000). Hunting success increases with the number of hunters, so that large groups in which all roles are performed are very successful (63 to 89% of captures achieved). During such collaborative hunts, each hunter synchronizes and spatially coordinates his movements to those performed by others, and sometimes anticipates their future actions. Each individual hunter can perform most complementary roles and individuals may even shift roles during a given hunt, demonstrating a capacity for role reversal and perspective taking. Tomasello et al. suggest that a chimpanzee hunter "simply assesses the state of the chase at each moment and decides what is best for it to do." However, drivers and ambushers achieve only 1% and 11% of the captures respectively, while 81% are achieved by individuals following the hunt from the ground. Consequently, drivers are granted about three times less meat than captors of the prey (Boesch 2002; Boesch & Boesch-Achermann 2000). Interestingly, ambushers that anticipates movements of the prey and the other hunters are granted an amount of meat equal to captors, even when they have not made the capture.

Thus, under a selfish hypothesis, chimpanzees should only wait on the ground for the prey to fall or perform the ambusher role that guarantees more meat. Group hunting would become rare. This is not the case as Taï chimpanzees hunt about 250 times per year (Boesch & Boesch-Achermann 2000). On the other hand, a joint goal hypothesis seems more compatible with the observations, with individual hunters assessing whatever role needs to be performed for the joint hunt and able to flexibly perform the roles needed independently of their short-term benefit. Like in a team of soccer players, individuals react opportunistically to the present situation while taking in account the shared goal of the team. Some players will rarely make a goal, like defenders and goalies, but the success of the team will critically depend upon their contribution. This is very reminiscent to group hunting in chimpanzees where synchronization of different coordinated roles, role reversal, and performance of less successful roles favor the realization of the joint goal. Thus, the group hunting behaviour of the Taï chimpanzees fulfills the criteria set by Tomasello et al. for shared goals and intentions. I am not claiming that chimpanzees perform like humans; I am merely emphasizing that the evidence published on hunting in chimpanzees is compatible with the scenario of shared goals and intentions proposed by Tomasello et al.

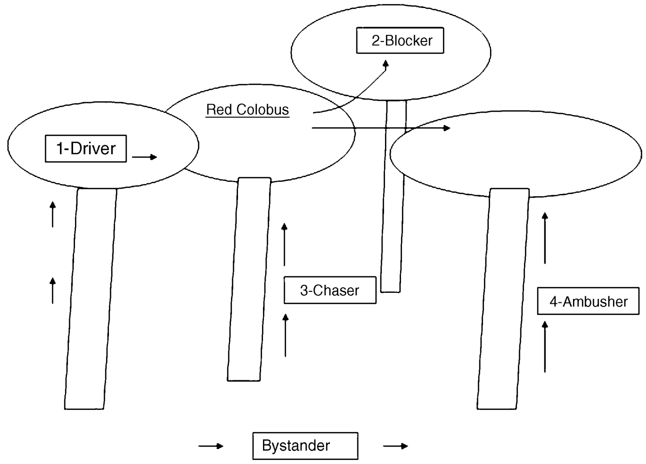

Figure 1 (Boesch). Illustration of a "typical" joint collaborative hunt in Taï chimpanzees indicating the spatial coordination of the different roles. The numbering indicates the approximate order in which the roles are joining into the hunt.

and therefore not a distinct human feature. One possible difference might be that human soccer players sometimes explicitly plan movements or strategies before the play starts and we have not yet seen this kind of shared planning in chimpanzees.

In the broader interest of the field of comparative psychology one further aspect is worth addressing: Why did Tomasello et al. ignore the published evidence on wild chimpanzee group hunting? Such an attitude is far from being isolated as illustrated by the conspicuous scarcity of reference to observations on wild animals in some of the cognitive literature claiming human superiority (e.g., Evans 2003; Heyes 1994; 1998; Povinelli 2000; Tomasello 1999). Generally, there is a tendency in comparative psychology to accept only experimental data. Observational data are dismissed as mere anecdotes or are discredited as not conclusive because alternative scenario could always been constructed. However, if we want to understand the specificity of cognitive abilities in humans and chimpanzees we have to take in account what they do in real life. Such data are irreplaceable as they provide the necessary information about how human and non-human primates perform. My point is not that field data answer all the questions about mental processes. What I am suggesting is that we need to formulate our hypothesis about human uniqueness in terms of performance that we should confront to the known performance of animals. The outcome could then be used as a guide for the aspects requiring more evidence, including experimental studies. Had that been done in Tomasello et al.'s article, I would probably have had no critical comment to forward on shortcomings or premature conclusions.

ACKNOWLEDGMENT
The author wants to thank Hedwige Boesch, Josep Call, and Peter Walsh for helpful comments.

Early development of shared intentionality with peers

Celia A. Brownell, Sara Nichols, and Margarita Svetlova
Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260.
brownell@pitt.edu srn1@pitt.edu pumita@yandex.ru

Abstract: In their account of the origins of human collaborative abilities, Tomasello et al. rely heavily on reasoning and evidence from adult–child collaborations. Peer collaborations are not discussed, but early peer collaborations differ from early adult–child collaborations. Describing and explaining the similarities and differences in shared intentionality with peers and adults will bring us closer to understanding the developmental mechanisms.

What are the origins of human collaborative abilities? Tomasello et al. hypothesize that humans possess a species-unique motivation to "feel and act and perceive together with others." This special motivation to share intentions, combined with intention understanding acquired in the context of adult–child collaborations, is proposed to drive the genesis of collaborative activity. By this account, at the end of the first year of life human infants are able to understand others' emotions, perceptions, intentions, goals, and plans. And, because they are uniquely motivated to share their psychological states with others – that is, to represent others' psychological states in concert with their own – human infants are able to collaborate with others and become, effectively, members of and contributors to human culture.

Although there is much to recommend this account, it depends exclusively on the role of adult–child collaborations. Nowhere are peer collaborations discussed. In our lab, we have studied early peer collaboration on tasks that require sharing a simple goal, and we find little evidence of either collaborative understanding or motivation to collaborate with peers until the close of the second year of life or well into the third year of life (Brownell & Carriger 1990, 1991; Brownell et al. 2003). Others have likewise suggested that collaborative peer play emerges toward the end of the second year of life (Asendorpf & Baudonniere 1993; Eckerman & Whitehead 1999; Eckerman et al. 1989). Only in the third year does cooperative play and communication with peers explicitly take into account the peer's actions, desires, and intentions (Ashley & Tomasello 1998; Smiley 2001).

These differences in development are not trivial. Such evidence does not call into question the assertion of Tomasello et al. that cultural cognition depends on shared intentionality, a point with which we fundamentally agree. But it does raise potential alternative developmental sequences and pathways, which in turn may introduce new explanatory demands and the possibility of other mechanisms. In particular, it suggests that shared intentionality may itself develop.

On the whole, infants and young toddlers do not appear particularly interested in social exchange with agemates, in contrast to their interest in collaborating socially with adults or even older siblings (Dunn 1988). Among 12-month-olds, familiar peers engage in simple social exchanges, such as looking and vocalizing to one another, less than once per hour. This increases over the second year to about once per five minutes at 24 months (Eckerman & Peterman 2001). Cooperative play with peers emerges between 20 and 24 months (Eckerman & Whitehead 1999; Eckerman et al. 1989; Howes 1988), and increases markedly between 24 and 28 months of age (Eckerman et al. 1989). Thus, the motivation to share intentions does not apply equally to all other persons early in development. One possibility for such a motivational difference is the developmental preeminence of attachment relationships during infancy (Brownell & Hazen 1999). Perhaps, in fact, it is their attachment relationships that make social engagement emotionally rewarding for infants and that first motivate them to share their emotions, desires and intentions.

Not only is children's interest in peer collaboration relatively late developing, but their ability to represent the peer's goals, in-