The relation between language and theory of mind in development and evolution*

Bertram F. Malle
University of Oregon

When Tom Givón asked me a while ago what my chapter would be about, I said, "Roughly, about the relation between language and theory of mind." His laconic response was, "Well, they are inseparable." So, I thought, there goes my chapter.

But not really. There is reason to believe that language and theory of mind have co-evolved, given their close relation in development and their tight connection in social behavior. However, they are clearly not inseparable — neurologically, cognitively, or functionally. So the question becomes, "What is the *exact* relation between language and theory of mind, in evolution, development, and social behavior?" To answer this question is a daunting task; I will try merely to clear a path toward an answer. I will consider several possible relations between the two faculties, bring conceptual arguments and empirical evidence to bear on them, and end up arguing for an escalation process in which language and theory of mind have fueled each other's evolution.

1. Some stipulations about language and theory of mind

1.1 Language

What constitutes genuine language? Modern language is typically described on several distinct levels: phonology, morphology, syntax, and semantics. To define language, the phonological level should be unimportant, because sign systems can be full-blown languages (Morford, this volume; Goldin-Meadow, this volume). Also, even though phonetic advances may have played an important role in the evolution of human speech (Studdert-Kennedy 2000b), very early forms of language may not have been spoken but signed (e.g., Corballis 2000; Donald 1998; Givón, this volume).

The next two levels, morphology and syntax, can vary extensively within systems that should count as language. I am not just referring to the difference

between analytic and synthetic languages but to the degree that a symbolic system includes (in morphology, syntax, or lexicon) central cognitive distinctions, such as perspective (I—you), modality (is—ought), spatial and temporal specifications, causality, intentionality, and so on. If a representational system contains none of these distinctions (e.g., a purely depictive system such as photography), it should not be called a language. A key component of language is that it offers choices in its representational repertoire. (Drawing may thus be called a language if it is understood to offer choices and distinctions). But the number of distinctions will vary, and so will the specific format used to mark these distinctions (such as morphology, syntax, or lexicon).

That a language must be a symbolic system (i.e., some entities within the system are understood to represent some entities outside the system) is uncontroversial. However, we must allow wide variation in semantic complexity. Just one notch outside language I would position expressive systems, such as unintentional screams of pain and satisfaction, facial features of emotion, and so on. Because they allow no choice in expression, they are not language (Li, this volume). They still "represent," but they do so by their very nature, not because a language user decides to use them this way.[1] (In Paul Grice's 1957, terms, they have *natural* meaning.) When symbols are iconic but under the control of the user, we have a primitive form of language. Ullin T. Place (2000) writes:

> Before human beings developed the ability to communicate vocally, they communicated with one another by means of some form of sign language whose primary function was to organize the cooperative social activities involved in hunting and foraging. In such a language the signs wold be entirely iconic, that is to say they would function as signs entirely by virtue of a resemblance between the sign and the object or movement it stands for.

Michael Corballis (2000) concurs:

> Language evolved from manual gestures, since gestures have at least the potential to represent concepts iconically rather than in abstract form. Once a set of iconic representations is established, increasing usage can then lead to more stylized and ultimately abstract representation, as has occurred in the evolution of writing systems.

As long as the iconic signs are understood to represent something else (an object, action, or relation), they are symbolic, and such a system of symbols, when used deliberately, should count as a primitive language. More elaborate systems will, for reasons of generativity, make use of arbitrary signs that represent a variety of entities, concrete or abstract, existent or nonexistent, sensory or conceptual.

Implicit in what I have said so far is an assumed feature of language that seems obvious once mentioned: Genuine language is shared among a community of speakers, including at least two (such as in the "secret" language of close childhood friends), but more likely involving a whole group (Wittgenstein 1953). Sometimes it is argued that the function of language in a social group is "for communication." However, this position pushes us only one step further back: What is communication for? I assume that communication–or better, language in communication–has as its major adaptive advantage *improved social coordination.* Such coordination has adaptive advantages in response to environmental challenges, in warfare, organized hunting (Whiten 1999), predatory defense, and social foraging (Bickerton 2000). Of course, it also enables teaching and learning (Tomasello 1996) and the strengthening of social bonds (Dunbar 1993). Whether these adaptive advantages of social coordination were in fact the driving forces for the emergence and refinement of language is a question we cannot answer until we consider other possible adaptive advantages, such as being a generative format for thinking and planning (Vaas 2000) or for reasoning about other minds (de Villiers 2000). I will have more to say about adaptive functions later.

1.2 Theory of mind

Theory of mind refers to the ability to represent, conceptualize, and reason about mental states. In its fully mature stage, theory of mind is a domain-specific conceptual framework that treats certain perceptual input as an agent, an intentional action, a belief, and so forth. This framework can be activated very easily, as Heider and Simmel (1944) have shown with stimuli as simple as triangles that move about in space. Theory of mind arguably underlies all conscious and unconscious cognition of human behavior (Malle 2001a), thus resembling a system of Kantian categories of social perception — i.e., the fundamental concepts by which people grasp social reality (Kant 1998/1787). But the framework not only classifies perceptual stimuli; it also directs further processing of the classified input, including inference, prediction, and explanation (Malle in press).

Of course, humans are not born with a fully mature theory of mind, so the developmental precursors of this theory are revealing components of the cognitive system that is at issue here. In particular, the conceptual framework of agency, intentionality, and mind most likely grows out of perceptual discrimination in infancy, between birth and about 18 months. All we see at birth is a capacity to imitate (Meltzoff and Moore 1977, 1989). But by about 9 months we find reliable evidence of children's perceptual sensitivity to self-propelled movements and to goal-directed action (Gergely, Nádasdy, Csibra, and Bíró 1995; Wellman and Philips 2001; Woodward 2001). By 14 months we see an ability to parse human action streams into meaningful (intention-relevant) units (Baldwin, Baird, Saylor, and Clark 2001; Sharon and Wynn 1998). And by 18 months we see an ability to infer intentions even from unsuccessful surface behavior. Out of this stage of perceptual discrimination and early concept formation grows the conceptual

understanding of desire (by age 2) and belief (by age 3), culminating in an understanding of false belief (by age 4), which many consider the watershed of theory of mind development (Perner 1991; Wellman 1990). At this point children understand that different people represent the world in different ways and therefore can believe or know different things. The concepts of this emerging framework are tightly interconnected, with the intentionality concept being an organizing node (Malle, Moses, and Baldwin 2001), and children show increasing versatility in their reasoning with these concepts (Baird 2000; Malle and Knobe 1997a).

The emergence of this conceptual and processing framework can be seen as an attempt to strike a balance between two counteracting forces: increasing self-other differentiation on the one hand and increasing self-other coordination on the other hand. After almost complete dependence in the early months of life, the child's autonomy increases over the next few years, and with it a potential for separation from the caregiver. Attachment researchers speak about the dangers of physical and emotional separation and postulate an evolutionary response in the form of an affective bond (Bowlby 1969). However, there is also the threat of mental separation, which can be illustrated most powerfully with the autistic child's disconnection from the caregiver's attention and emotion (Sigman, Kasari, Kwon, and Yirmiya 1992). As a result of such disconnection, observational and instructional learning, interpersonal coordination, and social information processing are all seriously hampered (Baron-Cohen 1995). With little sensitivity to others' mental life, one's own mental life may remain mysterious; and with little self-insight, mental interpretations of others' behaviors may be impossible (see Goldman 1989; Gordon 1986).

Among the functions of a theory of mind, the achievement of social coordination (of both behavior and mind) seems critical. In the course of development, the coordination of minds may become more important precisely when the coordination of behaviors becomes difficult, that is, when behavioral responses between self and other begin to diverge. Such divergence may come about, for example, because of differential motives and affect, multiple opportunities to act, and multiple interaction partners.

After these sketches of both language and theory of mind, I begin with my explorations of possible relations between the two.

2. Does one faculty precede the other?

2.1 Language precedes theory of mind

In adult social behavior, language is an important vehicle by which theory of mind skills are expressed and put to use. For example, behavior explanations, funda-

mental to making sense of other minds, are to a significant extent embedded in conversation (Kidd and Amabile 1981; Hilton 1990; Malle and Knobe 1997b) and exhibit complex linguistic regularities (Malle 1999; Malle, Knobe, O'Laughlin, Pearce, and Nelson 2000). Even though some causal judgments may be pre-verbal (Cheng and Novick 1990; van Overwalle 1998), those that explicitly take into account other minds seem to rely on a sophisticated conceptual framework that is unlikely to be pre-verbal (Malle 1999). Similarly, explicit mental state inferences (such as attempts to empathize with or take the perspective of another person) rely on verbal reasoning, representation, and eventually expression (e.g., Ickes 1993). Thus, to the extent that we understand theory of mind as a high-level skill of making sense of (others') mental states, language is almost invariably enmeshed in it (Smith 1996).

As argued earlier, however, restricting theory of mind capacities solely to high-level (and explicit/conscious) processes would be misleading. The developmental literature has identified a number of theory of mind skills that predate early language competence, such as classification of self-propelled movement (Premack 1990), eye-gaze tracking (Farroni, Johnson, Brockbank, and Simion 2000), imitation (Meltzoff and Moore 1977), recognizing goal-directed action (Gergely et al. 1995; Wellman and Philips 2001; Woodward 2001), and parsing the behavior stream into intention-relevant units (Baldwin, this volume; Baldwin et al. 2001). These skills do not incontestably demonstrate reasoning about *mind* (Povinelli 2001), but they are at least precursors or early forms of mental-state inference without which later, more explicit forms may well be impossible.

Perhaps the most striking evidence against the notion that language precedes theory of mind (at least in development) is that language acquisition itself appears to rely on theory of mind skills. Research demonstrates that the capacity to engage in joint attention (i.e., registering that self and other are both attending to the same object) is critical in early word learning and referential communication (Baldwin 1991, 1993; Tomasello 1998). Autistic children, by contrast, who have great difficulties with joint attention, show difficulties in language learning (Carpenter and Tomasello 2000), especially compared to children with other mental handicaps, who have general cognitive deficits but no deficiency in joint attention (Baron-Cohen, Baldwin, and Crowson 1997).

Another recent suggestion about theory of mind capacities involved in language comes from Abry and Labossiere (2000), who claim that vocalizations in communication require a monitoring system to distinguish own thought, own vocalization, other's vocalization, and other's thought. However, even though communication may indeed presuppose a self-other distinction, it does not require a thought-action distinction; for otherwise every organism's goal-directed action would presuppose this distinction and therefore every organism capable of goal-directed action would automatically have a theory of mind, which is obvi-

ously not the case. Early or primitive communication may well exist without consideration of the other's *mind*; it would merely expand the organism's attempts to influence the other's behavior by using certain (iconic) expressions.

2.2 Theory of mind precedes language

Some of the arguments against language preceding theory of mind (e.g., joint attention as a prerequisite of language acquisition) suggest that theory of mind may precede language (e.g., Baron-Cohen 1999). And indeed, several authors maintain that having the ability to infer others' communicative intentions is a precondition for language (e.g., Origgi and Sperber 2000; Sperber 2000). Suppose a social perceiver infers that an agent's expressive gesture G indicates a certain mental state M. Once the agent infers that the social perceiver made such a $G \rightarrow M$ inference, the agent can use G to intentionally communicate to the perceiver that M obtains. This simple scheme could be expanded from natural, iconic symbols to increasingly arbitrary symbols; and it can include a variety of referent mental states, such as bodily states, feelings, desires, plans, perceptions and perhaps even beliefs. Note that such a communicative system would have to be used primarily for coordination, because deception (the agent communicates G without M obtaining) would, over time, weaken the perceiver's willingness to make the $G \rightarrow M$ inference and thus weaken the semantic bond that people perceive between G and M.

Some scholars have gone one step further and argued that, in principle, once an organism uses a *symbol*, a convention is involved, requiring that the symbol user assume that others interpret the symbol the same way as the symbol user does (e.g., Hobson 2000). If one defines *symbol* in this way, and if language uses such symbols, then it follows that theory of mind precedes language. However, could an organism not use a symbol without an appreciation of minds and meanings? Suppose Franzy, a chimpanzee, displays a red card whenever he wants to play ball. Does he have to understand that the card signifies "I want to play ball" and understand that the caretaker knows that the card signifies 'Franzy wants to play ball'? Or could Franzy just have acquired a habit of prediction such that, whenever he has the urge to play ball, he shows the red card to his caretaker, because he has learned that under these circumstances his wish is likely to be fulfilled? In that case, would we say that Franzy communicates with his caretaker? If communication is defined in its functional role as facilitating social co-ordination, we should seriously consider calling Franzy's behavior communication. If, however, we define communication by means of symbol use, symbol use by means of appreciation of semiotic conventions, and the latter by means of representations of others' mental states, we cannot escape the conclusion that theory of mind precedes language.

What would weaken the position that theory of mind precedes language? For one, there is evidence that metarepresentation itself may be dependent on language. For example, Harris (1996) argues that the child's well-known transition from appreciating desires to appreciating both desires and beliefs comes with engagement in conversation. However, Bartsch and Wellman (1995) have evidence showing that children's conversations first feature desire verbs and then, after about a year, begin to feature belief verbs alongside desire verbs. This pattern renders conversation per se a less likely candidate for proffering a belief concept—for why would this concept not emerge early in the conversational practice but rather a full year later?

De Villiers (2000) offers another line of evidence for the notion that advancements in language engender advancements in theory of mind. She reports data showing that 3- to 4-year-old children's mastery, in comprehension and production, of oblique complement structures (e.g., "He said it was raining"; "She thought the bowl was empty")[2] predicts children's success at false-belief tasks, a conservative index of metarepresentational capacity (Perner 1991). One problem with these data is that language mastery was also presupposed in the false-belief tasks, which were all verbal. The predictive relation may therefore rely on the common linguistic demands in the two types of measures (oblique complements and false beliefs) rather than on a causal influence of linguistic advances on theory of mind capacities. It is conceivable that children acquire some false-belief understanding before they can couch it in linguistic structures of the sort needed in standard, verbal false-belief tasks. De Villiers' (2000) data would therefore have to be replicated with non-linguistic false-belief assessments to shore up the argument for linguistic influence on theory of mind development.

But even if we interpret De Villiers's (2000) results as suggesting a leading influence of language competence on theory of mind competence, we are considering here fairly mature levels of language as well as theory of mind. The findings on earlier stages of development remain, which showed that theory of mind (ToM) skills are more likely to precede than to follow language skills. Thus, at best we can conclude that the direction of influence shifts from ToM $\rightarrow$ Language in early stages of development to Language $\rightarrow$ ToM in later stages. Pushing this claim further we would expect to see an increasing influence of linguistic mastery and verbal intelligence on the refinement of theory of mind skills in later childhood and early adolescence. Unfortunately, the literature on developing theory of mind past the age of 7 is very limited, so data of the sort needed have yet to be collected. However, what supports the suggestion that language becomes more autonomous from theory of mind in later stages of development is the fact that at least some autistic children are able to catch up on their language skills in later years without overcoming their deficit in theory of mind (Baron-Cohen 1995). This linguistic competence in turn allows them to compensate for their inability

to infer other people's mental states by using complex linguistic representations (multi-layered if-then rules) when interacting with others (Sacks 1995).

3. Did language and theory of mind co-evolve?

In light of the current developmental evidence, neither the position that language precedes theory of mind nor the position that theory of mind precedes language is compelling, so the claim that the two faculties co-evolved becomes increasingly plausible. However, claims of co-development and co-evolution are easily made; the challenge is to flesh out the details of such a claim and say exactly how and why the two faculties co-emerged. Moreover, treating language and theory of mind as monolithic entities will not do. A plausible model of co-evolution will have to include the specification of elementary components of both language and theory of mind, because it is likely that those components influenced one another in a gradual escalation.

But before I discuss such escalation, I would like to address the reader's possible discomfort with my quick transition from developmental to evolutionary conclusions about the possible ordering of language and theory of mind.

There are two questions to consider, of which the first one is this: Even if language does not precede theory of mind in development, could it nevertheless precede it in evolution? I would claim that this is highly unlikely, because children in the evolutionary past had to learn their language too, and we know of no other plausible mechanisms for acquiring language than through social-interactive learning, imitation, and practice, relying heavily on early theory of mind skills. What still seems feasible is that some sort of protolanguage could have been acquired without a theory of mind (see Franzy's symbolic use under section 1.2), whereas full-fledged language became possible precisely because of the evolution of a theory of mind. I will return to this chain model of *protolanguage* → *theory of mind* → *language* in the next section.

The second question is this: Even if theory of mind did not precede language in development, could it nevertheless precede it in evolution? This seems quite plausible. First, as we have seen, any documented influence of language on theory of mind occurs later in development and is therefore compatible with an evolutionary primacy of theory of mind. Second, to the extent that only advanced language skills can exert a causal influence on theory of mind progression, we still have to explain the evolutionary emergence of those advanced language skills, which seem difficult to acquire without some form of theory of mind.

These deliberations underscore the need for a fine-grained analysis of the multi-stage evolution of language and the possible multi-stage co-evolution of theory of mind. I organize my thoughts on such an analysis into three sections: Co-evolution

with coincidental parallelism; co-evolution due to the emergence of a third factor that enabled both language and theory of mind; and co-evolution in the form of mutual escalation toward a shared, or at least partially shared, adaptive advantage.

3.1 Coincidental concurrent evolution

The weakest model of co-evolution would assume no shared adaptive pressures and no mutual influence during evolution. The two faculties would then have evolved for distinct reasons, though roughly during the same time period. According to this model, humans just happen to have both a linguistic capacity and a refined theory of mind, and their respective architectures would have been established independently, leading to no more than coincidental functional parallels between the two.

Several facts speak against this position. First, we find intimate ties between language and theory of mind both in development (e.g., language learning relies on theory of mind) and in adult social behavior (mental-state inferences are required and practiced in conversation). Second, there is likely a common adaptive advantage to both faculties in the form of vastly improved social co-ordination (which does not rule out additional, unshared adaptive advantages, such as art, planning, counterfactuals, etc.). Third, the striking parallels between fundamental linguistic structures (e.g., thematic roles, verb classes, and perspective) and their corresponding elements in theory of mind (the concepts of agency, intentionality, self vs. other; see Malle 2001b) would remain unexplained if we did not postulate an interdependence between the two faculties during evolution. Such an interdependence can be treated like a correlation between variables—it may be due to a causal influence of one on the other or due to a third factor that exerts a causal influence on both. Below I explore the possibility of a third factor and then turn to the scenario of mutual causal influence.

3.2 In search of a third factor

Two types of factors might have fostered the co-evolution of language and theory of mind: a powerful adaptive demand or another adaptation that made both faculties possible. As candidates for the latter we may think of adaptations such as consciousness, executive control, increased memory capacity, representational ability, or propositional thought. The problem with all of these factors is that they are at least as likely to be consequences of language and theory of mind as they are antecedents of them (see Bogdan 2000, and Carruthers 1998). More important, neither theory of mind nor language necessarily requires the operation of these factors. For example, theory of mind is not particularly memory-intensive and many of its functions do not require consciousness or executive control (Malle 2001a). Similarly, phonology and syntax do not require consciousness or executive

control (although syntax may require fairly sophisticated memory). In addition, all of the mentioned third factors leave most of the specific features of language unaccounted for (e.g., Hurford 2000). The same may hold true for features of theory of mind. The concept of intentionality, for example, can in no way be derived from any cognitive mechanisms in consciousness, memory, or the like. Concepts such as intentionality must be adaptations to a reality (or a conceived reality) that is better managed and mastered with than without these concepts. Of course, advances in cognitive processes may improve the resolution, efficiency, and manipulability of a conceptual framework, but the concepts themselves must be there in the first place, as an adaptation to the social world around.

This leads us to the second potential driving force for the co-evolution of language and theory of mind: a consistent external demand that had an impact on both faculties. The most likely candidate for such a demand is the need for improved social coordination, with at least two things speaking for it. First, a number of different environmental challenges that early hominids encountered (e.g., climate shifts, wildlife disruptions, novel germs, increased group size, migration, hostile encounters with other hominid groups) likely resulted in a need for greater social coordination. When food is tight, life is novel or dangerous, and groups become large, social coordination and organization are called for. Second, among the contemporary functions of language and theory of mind, improved social coordination is the most salient function they share. (I am not aware of any other shared functions; but I would be happy to be corrected.)

By itself, a powerful demand such as social coordination would not explain the emergence of two adaptations. However, a plausible explanation can be constructed if we presume that one of the faculties emerged first in a primitive form and facilitated a primitive form of the other, and soon the two plunged into a race in which advances in one faculty repeatedly enabled or demanded advances in the other. What we have here is a version of the arms-race argument, but I am thinking more of an arms race between psychological faculties than between different genetic pools, and so I prefer the term *escalation*.

4.   Evolutionary escalation

The balance of evidence reviewed earlier favors the emergence of a primitive theory of mind (TOM-1) as a first adaptive step, because language learning is rather inconceivable without it. What might such a primitive theory look like? Povinelli (2001) reviewed the literature to date on primate theory of mind and concluded that chimpanzees have a refined behavior reading and detection system but no clear understanding of the mind as the underlying source of observed behavioral patterns. Unenculturated chimpanzees also show no reliable skills of

joint attention (Tomasello 1996), and the evidence on imitation is mixed (Smith 1996; Whiten 1999). Thus I postulate that TOM-1 includes:

a.   a capacity for imitation (a mechanism that matches others' behavior to self's internal states and provides a necessary condition for inferences about those internal states in others; Meltzoff and Brooks 2001);
b.   a capacity for joint attention (representing oneself and another person attending to an object); and
c.   an inferential sensitivity to some mental states associated with action (approximating a desire or goal concept).

Most likely, the three elements emerged in the listed order and built on each other until reaching a stage that is roughly comparable to a modern 9–12 month-old infant, which is also the stage when language learning begins. In human evolution, however, the emergence of these three components may well have been separated by a few million years (MacWhinney, this volume).

Imitation and joint attention may be easy postulates to accept, but how could a sensitivity to desires have emerged? The early hominid brain may have fed data from behavioral pattern analysis into a reasoning device that could make inferences about unobservable forces. Some have argued that these inferences are enabled by a sort of analogical reasoning from one's own case (Goldman 2001; Meltzoff and Brooks 2001). This system would match up patterns in the other's behavior with patterns in one's own behavior (Blakemore and Decety 2001), then patterns in one's own behavior with (a vague awareness of) one's own mental state, finally leading to a projective inference of one's own mental state to the other person (Gallese and Goldman 1998). Such an analogical process has few rivals in explaining what pushed the primate perceiver from sophisticated behavior analysis to the crude beginnings of a mental analysis. If this notion of an analogical process hits the mark, however, then a mature TOM-1 would presuppose advances in introspective access (as vague as it may be), and such introspective access may qualify as a "driving third variable." How such introspective access might have evolved is unclear. Some form of brain reorganization seems inevitable (Humphrey 2000), perhaps with increases in self-other differentiation playing a crucial triggering role. For example, encounters with variants of one's genus or species would highlight how the self is different from the other, and living in larger-sized groups might expand the gene pool to allow for more individual differences, leading to more self-other distinctions. In both cases, global awareness of self as different from others could set the stage for more specific awareness of self's distinct preferences, emotions, and so on.

If we equip TOM-1 with the three elements of imitation, joint attention, and inferential sensitivity to desires, these elements would likely reinforce one another: Joint attention facilitates sensitivity to others' object-directed desires, and desire

sensitivity improves imitation of behaviors as well as of goal states (Meltzoff 1995). Operating together, the three elements also open new possibilities. Imitation and joint attention may initiate cooperative learning, especially if the learner also recognizes the teacher's goal in the process. Joint attention with desire sensitivity should improve behavior prediction and thus the coordination of shared activities (e.g., hunting, combat, tool making). And, most important, the three elements would enable expressive gestures or vocalizations to become predictably associated with certain objects (especially objects of goal-directed actions), laying the groundwork for a simple protolanguage (Bickerton 1990). Specifically, imitation would enable cross-member reliability in vocalization (De Boer 2000; Studdert-Kennedy 2000a); joint attention would enable cross-member reliability for acts of vocal reference (Baldwin 1993); and desire sensitivity would launch the interpretation of vocalized reference as a form of communicative intention.

At the next stage of the hypothesized escalation process, the achieved communication system becomes a spring board for further advances in mental-state inference. We can assume that hominid agents' mental attitudes were more complex than could be represented by contemporary social perceivers' mental inferences. The expanding vocabulary of protolanguage may then have taken on an intermediary role by differentiating between various relations that agents might have had toward objects (e.g., like vs. dislike; weak vs. strong desire; desire vs. intention) and by differentiating between types of behaviors–primarily intentional vs. unintentional, but also among specific speech acts such as command, request, offer, decline, or warning. If used reliably, and practiced over numerous generations, these communicative tools must have refined early hominids' conceptual framework of mind and behavior.

But there was likely another process at work that helped refine theory of mind. The increasing vocabulary and differentiation of this early communication system introduced a great deal of ambiguity and inferential challenge. For one thing, the linguistic forms included in protolanguage were probably not names for objects (Wray 2000). Instead, they were speech acts (commands, threats, warnings etc.) that holistically expressed person-object relations and perhaps included interjections to clarify such person-object relations (e.g., a prehistoric "oops" to indicate unintentionality or an ancient "there" to indicate intentionality; see Carpenter, Akhtar, and Tomasello 1998). Add to that the fact that even in modern times communicators overestimate the clarity of their expressions and the transparency of their communicative intentions (Gilovich, Savitsky, and Medvec 1998; Keysar 1994). We can conclude, then, that protolanguage posed serious difficulties for the ancient social perceiver and may have caused a fair amount of misunderstanding and conflict within and between groups.

In response to both the opportunities and challenges posed by protolanguage, theory of mind must have advanced to TOM-2, through an extension of intro-

spective access and/or an improvement of cognitive capacities, such as inference, conceptual distinctions, contextual awareness, and memory. The change in memory capacity may be a particularly important link, because protolanguage should have increased interactions among group members (in learning, teaching, and joint action), which in turn demanded better memory for one's relations with each group member, especially if the groups also increased in size. One may think here of cheater detection and tracking the balance of one's reciprocal exchanges (Cosmides and Tooby 1992). But the tracking must have gone further, to improved representations of people's habitual desires, skills, and behavior trends. Reliable memory representations of interactions with others would also allow the social agent to better recognize individual differences among people in their desires, skills, and attitudes toward the agent. These emerging differentiations may then be the starting point for the realization that people represent the world in different ways and act on those representations (Perner 1991).

Once mental-state inferences in TOM-2 were sharpened and memory for people and interaction contexts was recruited to disambiguate communication, language itself progressed toward greater clarity. This may seem paradoxical, because it was precisely the ambiguity of protolanguage that demanded the more refined TOM-2, and one might expect that this refinement would have allowed protolanguage to remain ambiguous. But I think otherwise. Consider a conversation between two people in which one tries to express a complex internal state and is rather vague and incoherent. A listener who genuinely wants to understand the speaker will pose questions to clarify what the speaker means, will propose and "test" various formulations, until the speaker feels that a certain formulation adequately depicts his or her internal state. Thus, the two negotiate and agree upon a set of expressions that the speaker will use henceforth to describe the relevant internal state (see Krauss and Fussell 1991). Thus, modern speakers become clearer and more precise in their expressions when listeners use their mental-state inference skills to help disambiguate expressions, and I would assume that something similar occurred in the evolutionary past.

These processes must have given rise to a more advanced protolanguage (PL-2), which was less holistic and more precise and incorporated a greater number of linguistic references to internal states. Some basic grammatical distinctions may have been added, such as between names for objects (nouns) and names for activities (verbs). Increasing intragroup interactions may have also demanded names for people, and combinations between these three classes of words would have permitted depiction of important social events, such as a certain person manipulating a certain object.

If PL-2 was spoken by adults around young children, its representational uses should have helped children differentiate among their own emerging behavior patterns, their preference structures, and even their social emotions, which would

in turn sharpen their introspection from an earlier age on. Thus we can derive the modern-day prediction that the degree to which children are exposed to representational language should accelerate their introspective abilities. At the opposite end would be deaf children with parents whose sign language is primitive and lacks representational elaboration, resulting in slowed maturation of the children's introspective abilities.

Sharpened introspection within any new generation may then lead to increasingly subtle and differentiated first-person statements about preferences or emotional responses, combining the functions of mere expression in protolanguage-1 with those of representation in protolanguage-2. Successive generations would thus refine the shared linguistic tools, not unlike successive generations of signers have been shown to refine their language (Senghas 2000). Once more, these advances have implications for theory of mind, because if agents make their internal states explicit to others (at least under conditions of intimacy and trust), social perceivers receive more validating information for their mental state inferences, which would become increasingly differentiated and spiral up toward a level of TOM-3.

How and why fully syntactic language emerged out of PL-2 and how TOM-3 differs from a full-fledged modern theory of mind would have to be the topic of another paper. The same holds for the intriguing idea that contemporary scientific psychology may be the latest step in this escalation between language and theory of mind, because it tries to offer explicit verbal representations and extensions of folk psychology (theory of mind). That then raises the question whether, on the next level of the spiral, theory of mind will change under the weight of scientific findings, as some have confidently predicted (Churchland 1988). Whatever future evolution may bring, the evidence at hand suggests that language and theory of mind evolved in grades (MacWhinney, this volume) and in constant interaction, serving one primary adaptive goal: to improve social coordination.

5.   Coda

Dan Slobin, skeptical about the scientific study of language evolution (Slobin, this volume) characterized my preceding proposal as an excellent science-fiction story. Being rather skeptical myself of much of evolutionary psychology, I tend to agree with this characterization. But an open-minded view of science (finally shedding the shackles of neopositivism) allows for speculative, even fictional models to serve a heuristic or hypothesis-generating function. A cognitive-evolutionary model of what happened, say, 100,000 years ago will never be empirically testable in the strict sense; but neither will an astrophysical model of what happened 100 Million years ago. In both cases we build models that best fit the facts we have available, and we often begin to search for new facts that would have to be true if the model were

correct. It may take a while to find these facts (or their refutation), and several competing models may be "best-fitting." But no philosophy of science should prevent us from discussing, comparing, and challenging these models. In this spirit, the skeptical reader as well as the aficionado of science fiction are invited to propose a better model than the one here offered, point to existing facts overlooked, and search for new facts not yet considered. It is this method of scrutinizing claims and models in a rational, systematic manner that counts as science.

Notes

1.   If an organism has control over expressive features and begins to conceal or feign them for particular purposes, I would still not call it a language but rather a sophisticated manipulation of the other organism's perception or (more likely) behavior. Whether or not one believes the arguments that some animals are capable of full-blown deception (manipulating the other's mind, not just behavior) is irrelevant here, because language requires at least two people who both understand that a symbolic system is being used.

2.   What is important about these structures is that the complement can be false. That is, the speaker understands that, even though the agent may have said or believed or wanted that it was raining outside, it may in reality not have been raining outside. This linguistic structure thus enables the representation of another person's subjective representation of reality, or "metarepresentation."

References

Abry, and Laboissiere, R. 2000. Who's afraid of the co-evolution of medial and lateral cortices for speech? Paper presented at the 3rd Conference on the evolution of language. Ecole Nationale Supérieure des Télécommunications, Paris, France. http://www.infres.enst.fr/confs/evolang/actes/_actes01.html.

Baird, J. A. 2000. *Young Children's Understanding of the Relation Between Actions and Intentions*. Dissertation, University of Oregon.

Baldwin, D. A. 1991. "Infants' contribution to the achievement of joint reference". *Child Development* 62: 875–90.

Baldwin, D. A. 1993. "Early referential understanding: Infants' ability to recognize referential acts for what they are". *Developmental Psychology* 29: 832–43.

Baldwin, D. A, and Tomasello, M. 1998. "Word learning: A window on early pragmatic understanding". In *Proceedings of the 29th Annual Child Language Research Forum*, E. V. Clark (ed), 3–23. Stanford, Calif.: Center for the Study of Language and Information.

Baldwin, D.A., Baird, J.A., Saylor, M.M., and Clark, M.A. 2001. "Infants parse dynamic action". *Child Development* 72: 708–17.

Baron-Cohen, S. 1999. "The evolution of a theory of mind". In *The Descent of Mind: Psychological Perspectives on Hominid Evolution*, M.C. Corballis and S.E.G. Lea (eds), 261–77. New York: Oxford University Press.

Baron-Cohen, S. 1995. *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, Mass.: MIT Press.

Baron-Cohen, S., Baldwin, D.A, and Crowson, M. 1997. "Do children with autism use the speaker's direction of gaze strategy to crack the code of language?" *Child Development* 68: 48–57.

Bartsch, K., and Wellman, H.M. 1995. *Children Talk about the Mind*. New York: Oxford University Press.

Bickerton, D. 1990. *Language and Species*. Chicago: University of Chicago Press

Bickerton, D. 2000. Foraging versus social intelligence in the evolution of protolanguage. Paper presented at the 3rd Conference on the evolution of language. Ecole Nationale Supérieure des Télécommunications. Paris, France. http://www.infres.enst.fr/confs/evolang/actes/_actes05.html.

Blakemore, S. and Decety, J. 2001. "From the perception of action to the understanding of intention". *Nature Reviews Neuroscience* 2: 561–7.

Bogdan, R.J. 2000. *Minding Minds: Evolving a Reflexive Mind by Interpreting Others*. Cambridge, Mass.: MIT Press.

Bowlby, J. 1969. *Attachment and Loss*. New York: Basic Books.

Carpenter, M., Akhtar, N., and Tomasello, M. 1998. "Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions". *Infant Behavior and Development* 21: 315–30.

Carpenter, M., and Tomasello, M. 2000. "Joint attention, cultural learning, and language acquisition: Implications for children with autism". In *Autism Spectrum Disorders: A Transactional Developmental Perspective*, A.M. Wetherby and B.M. Prizant (eds), 31–54. Baltimore: Paul H. Brookes.

Carruthers, P. 1998. "Thinking in language? Evolution and a modularist possibility". In *Language and Thought: Interdisciplinary Themes*, P. Carruthers and J. Boucher (eds), 94–119. Cambridge: Cambridge University Press.

Cheng, P.W., and Novick, L.R. 1990. "A probabilistic contrast model of causal induction". *Journal of Personality and Social Psychology* 58: 545–67.

Churchland, P.M. 1988. *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind* (Rev. edition). Cambridge, Mass.: MIT Press.

Cosmides, L., and Tooby, J. 1992. "Cognitive adaptations for social exchange". In *The adapted mind: Evolutionary psychology and the generation of culture*, J.H. Barkow, L. Cosmides, and J. Tooby (eds), 163–228. New York: Oxford University Press.

Deacon, T. 1997. *The Symbolic Species: The Co-evolution of Language and the Brain*. New York: W.W. Norton.

de Boer, B. 2000. "Emergence of sound systems through self-organization". In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, C. Knight, M. Studdert-Kennedy, and J.R. Hurford (eds), 177–98. Cambridge: Cambridge University Press.

De Villiers, J. 2000. "Language and theory of mind: What are the developmental relationships?" In *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*, S. Baron-Cohen, H. Tager-Flusberg, and D.J. Cohen (eds), 83–123. 2nd edition. New York: Oxford University Press.

Donald, M. 1998. "Mimesis and the excessive suite: Missing links in language evolution". In *Approaches to the Evolution of Language: Social and Cognitive Bases*, J.R. Hurford, M. Studdert-Kennedy, and C. Knight (eds), 44–67. Cambridge: Cambridge University Press.

Farroni, T., Johnson, M.H, Brockbank, M., and Simion, F. 2000. "Infants' use of gaze direction to cue attention: The importance of perceived motion". *Visual Cognition* 7: 705–18.

Gallese, V., and Goldman, A. 1998. "Mirror neurons and the simulation theory of mind-reading". *Trends in Cognitive Sciences* 2: 493–501.

Gergely, G., Nádasdy, Z., Csibra, G., and Bíró, S. 1995. "Taking the intentional stance at 12 months of age". *Cognition* 56: 165–93.

Gilovich, T., Savitsky, K., and Medvec, V.H. 1998. "The illusion of transparency: Biased assessments of others' ability to read one's emotional states". *Journal of Personality and Social Psychology* 75: 332–46.

Goldman, A.I. 1989. "Interpretation psychologized". *Mind and Language* 4: 161–85.

Goldman, A.I. 2001. "Desire, intention, and the simulation theory". In *Intentions and Intentionality: Foundations of Social Cognition*, B.F. Malle, L.J. Moses, and D.A. Baldwin (eds), 207–25. Cambridge, Mass.: MIT Press.

Gordon, R.M. 1986. "Folk psychology as simulation". *Mind and Language* 1: 158–71.

Grice, H.P. 1957. "Meaning". *Philosophical Review* 64: 377–88.

Harris, P. 1996. "Desires, beliefs, and language". In *Theories of Theories of Mind*, P. Carruthers and P.K. Smith (eds), 200–20. Cambridge: Cambridge University Press.

Heider, F., and Simmel, M. 1944. "An experimental study of apparent behavior". *American Journal of Psychology* 57: 243–59.

Hilton, D.J. 1990. "Conversational processes and causal explanation". *Psychological Bulletin* 107: 65–81

Hobson, R.P. 2000. "The grounding of symbols: A social-developmental account". In *Children's Reasoning and the Mind*, P. Mitchell and K.J. Riggs (eds), 11–35. Hove: Psychology Press.

Humphrey, N. 2000. "The privatization of sensation". In *The Evolution of Cognition*, C. Heyes and L. Huber (eds), 241–52. Cambridge, Mass.: MIT Press.

Hurford, J.R. 2000. The roles of communication and representation in language evolution. Paper presented at the 3rd Conference on the evolution of language. Ecole Nationale Supérieure des Télécommunications. Paris, France. http://www.infres.enst.fr/confs/evolang/actes/_actes37.html.

Ickes, W. 1993. "Empathic accuracy". *Journal of Personality* 61: 587–610.

Kant, I. 1998/1787. *Critique of Pure Reason*. (Translated and edited by Paul Guyer, Allen W. Wood). Cambridge: Cambridge University Press.

Keysar, B. 1994. "The illusory transparency of intention: Linguistic perspective taking in text". *Cognitive Psychology* 26: 165–208.

Kidd, R. F., Amabile, T. M. 1981. "Causal explanations in social interaction: Some dialogues on dialogue". In *New directions in attribution research*, Volume III, J. H. Harvey, W. J. Ickes, and R. F. Kidd (eds), 307–28. Hillsdale, NJ: Lawrence Erlbaum Associates.

Krauss, R. M., and Fussell, S. R. 1991. "Constructing shared communicative environments". In *Perspectives on Socially Shared Cognition*, L. B. Resnick, J. M. Levine, and S. D. Teasley (eds), 172–200. Washington: American Psychological Association.

Malle, B. F. 1999. "How people explain behavior: A new theoretical framework". *Personality and Social Psychology Review* 3: 23–48.

Malle, B. F. In press. "The folk theory of mind: Conceptual foundations for social cognition". In *The New Unconscious*, R. Hassin, J. S. Uleman, and J. A. Bargh (eds). New York: Oxford University Press.

Malle, B. F. 2002. "Verbs of interpersonal causality and the folk theory of mind and behavior". In *The Grammar of Causation and Interpersonal Manipulation*, M. Shibatani (ed), 57–83. Amsterdam: Benjamins.

Malle, B. F., and Knobe, J. 1997a. "The folk concept of intentionality". *Journal of Experimental Social Psychology* 33: 101–21.

Malle, B. F., and Knobe, J. 1997b. "Which behaviors do people explain? A basic actor-observer asymmetry". *Journal of Personality and Social Psychology* 72: 288–304.

Malle, B. F., Knobe, J., O'Laughlin, M., Pearce, G. E., and Nelson, S. E. 2000. "Conceptual structure and social functions of behavior explanations: Beyond person–situation attributions". *Journal of Personality and Social Psychology* 79: 309–26.

Meltzoff, A. N. 1995. "Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children". *Developmental Psychology* 31: 838–50.

Meltzoff, A. N., and Brooks, R. 2001. " 'Like me' as a building block for understanding other minds: Bodily acts, attention, and intention". In *Intentions and Intentionality: Foundations of Social Cognition*, B. F. Malle, L. J. Moses, and D. A. Baldwin (eds), 171–91. Cambridge, Mass.: MIT Press.

Meltzoff, A. N., and Moore, M. K. 1977. "Imitation of facial and manual gestures by human neonates". *Science* 198: 75–8.

Meltzoff, A. N., and Moore, M. K. 1989. "Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms". *Developmental Psychology* 25: 954–62.

Mithen, S. 1996. *The Prehistory of the Mind*. London: Thames and Hudson.

Origgi, G. and Sperber, D. 2000. "Evolution, communication and the proper function of language". In *Evolution and the Human Mind: Modularity, Language and Meta-cognition*, P. Carruthers and A. Chamberlain (eds), 140–69. Cambridge: Cambridge University Press.

Perner, J. 1991. *Understanding the Representational Mind*. Cambridge, Mass.: MIT Press.

Povinelli, D. M. 2001. "On the possibilities of detecting intentions prior to understanding them". In *Intentions and intentionality: Foundations of social cognition*, B. F. Malle, L. J. Moses, and D. A. Baldwin (eds), 225–48. Cambridge, Mass.: MIT Press.

Premack, D. 1990. "The infant's theory of self-propelled objects". *Cognition* 36: 1–16.

Sacks, O. 1995. "An anthropologist on Mars". In *An anthropologist on Mars*, O. Sacks, 244–96. New York: Vintage Books.

Senghas, A. 2000 Evolution of Grammar: The Grammaticization of Nicaraguan Sign Language by sequential iterations of native acquisition. Paper presented at the 3rd Conference on the evolution of language. Ecole Nationale Supérieure des Télécommunications. Paris, France. http://www.infres.enst.fr/confs/evolang/actes/_actes64.html.

Sharon, T., and Wynn, K. 1998. "Individuation of actions from continuous motion". *Psychological Science* 9: 357–62.

Sigman, M. D, Kasari, C., Kwon, J., and Yirmiya, N. 1992. "Responses to the negative emotions of others by autistic, mentally retarded, and normal children". *Child Development* 63: 796–807.

Smith, P. K. 1996. "Language and the evolution of mind-reading". In *Theories of Theories of Mind*, P. Carruthers and P. K. Smith (eds), 344–54. Cambridge: Cambridge University Press.

Sperber, D. 2000. "Metarepresentations in an evolutionary perspective". In *Metarepresentations: A Multidisciplinary Perspective*, D. Sperber (ed), 117–37. New York: Oxford University Press.

Studdert-Kennedy, M. 2000a. "Evolutionary implications of the particulate principle: Imitation and the dissociation of phonetic form from semantic function". In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, C. Knight, M. Studdert-Kennedy, and J. R. Hurford (eds), 161–76. Cambridge: Cambridge University Press.

Studdert-Kennedy, M. 2000b. "Introduction: The emergence of phonetic structure". In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, C. Knight, M. Studdert-Kennedy, and J. R. Hurford (eds), 123–9. Cambridge: Cambridge University Press.

Tomasello, M. 1996. "The cultural roots of language". In *Communicating Meaning: The Evolution and Development of Language*, B. M. Velichkovsky and D. M. Rumbaugh (eds), 275–307. Mahwah, NJ: Lawrence Erlbaum Associates.

Tomasello, M. 1998. "Reference: Intending that Others Jointly Attend". *Pragmatics and Cognition* 6: 229–43.

Vaas, R. 2000. Evolving language, I-consciousness and free will. Paper presented at the 3rd Conference on the evolution of language. Ecole Nationale Supérieure des Télécommuniycations. Paris, France. http://www.infres.enst.fr/confs/evolang/actes/_actes73.html.

van Overwalle, F. 1998. "Causal explanation as constraint satisfaction: A critique and a feed-forward connectionist alternative". *Journal of Personality and Social Psychology* 74: 312–28.

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

Wellman, H. W., and Phillips, A. T. 2001. "Developing intentional understandings". In *Intentions and Intentionality: Foundations of Social Cognition*, B. F. Malle, L. J. Moses, and D. A. Baldwin (eds), 125–48. Cambridge, Mass.: MIT Press.

Whiten, A. 1999. "The evolution of deep social mind in humans". In *The descent of mind: Psychological Perspectives on Hominid Evolution*, M. C. Corballis and S. E. G. Lea (eds), 173–93. New York: Oxford University Press.

Wittgenstein, L. 1953. *Philosophical Investigations.* (Transl. G. E. M. Anscombe). Malden, Mass.: Blackwell.

Woodward, A. L., Sommervile, J. A., and Guajardo, J. J. 2001. "How infants make sense of intentional action". In *Intentions and Intentionality: Foundations of Social Cognition*, B. F. Malle, L. J. Moses, and D. A. Baldwin (eds), 149–70. Cambridge, Mass.: MIT Press.

Wray, A. 2000. "Holistic utterances in protolanguage: The link from primates to humans". In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, C. Knight, M. Studdert-Kennedy, and J. R. Hurford (eds), 285–302. Cambridge: Cambridge University Press.