RADU J. BOGDAN

DEVELOPING MENTAL ABILITIES BY REPRESENTING INTENTIONALITY

ABSTRACT. Communication by shared meaning, the mastery of word semantics, metarepresentation and metamentation are mental abilities, uniquely human, that share a sense of intentionality or reference. The latter is developed by a naive psychology or interpretation – a competence dedicated to representing intentional relations between conspecifics and the world. The idea that interpretation builds new mental abilities around a sense of reference is based on three lines of analysis – conceptual, psychological and evolutionary. The conceptual analysis reveals that a sense of reference is at the heart of the abilities in question. Psychological data track tight developmental correlations between interpretation and the abilities it designs. Finally, an evolutionary hypothesis looks at why interpretation designed those new abilities around a sense of reference.

1. INTRODUCTION

Gestures, words, and often thoughts refer because people intend them to; and people can do that because they have a sense of the reference relation itself. It is a *generic* sense of reference that is more basic, broader and acquired earlier than understanding word or belief reference. This generic sense of reference is not merely the result of having a mind that represents the world. A mind that represents the world can be said to be *intrinsically* intentional: its states are about items and conditions in the world. Most animal minds are intrinsically intentional. Yet most animal minds do not *recognize* intentionality, aboutness or reference – relations that I treat as equivalent in this paper. Having beliefs or other mind-world relations is not the same as recognizing beliefs or other mind-world relations and need not lead to such a recognition. Evolution concurs. Very few kinds of minds, probably only the primate ones, recognize some aspects of intentionality and only human minds recognize full intentionality and thus develop a generic sense of reference. Only the latter minds come to know that words refer, beliefs represent and thoughts can be about other thoughts. Even though these forms of knowledge are different and develop at distinct stages, they seem to be built around the same generic sense of reference.

Synthese **129:** 233–258, 2001.

Many philosophers, psychologists and linguists look for the source of this amazing accomplishment in how the individual mind is designed to perceive and act on the physical world. They expect the sense of reference somehow to emerge from the mind's intrinsic intentionality. Social interaction is thought to be just an occasion and stimulus. A different and increasingly influential tradition, with roots in the works of George Mead and Lev Vygotsky from the first decades of 20th century, takes social interaction to be essential to a sense of reference. Its central hypothesis is that gestures, symbols, words and often thoughts refer because people intend them to refer, and people intend them to refer because they interact with and influence each other. The reference relation is an instrument that must be represented mentally to do its job. This line of analysis was explored in some early works in social psychology (Werner and Kaplan 1963) and psycholinguistics (Bates 1976) and echoed, in a more abstract way, in the philosophical work of Paul Grice (1957).

Yet still unexplained in these works was the very ability to represent reference. As a result, still unanswered were questions about the origin and rationale of the generic sense of reference and its impact on the minds that acquire it. That began to change about 20 years ago, thanks to a spectacular burst of interest, a rapidly growing body of research and some remarkable results in the primate and developmental study of naive psychology or theory of mind, as psychologists call it, or folk or commonsense psychology, according to philosophers. My terminological preference is for interpreting minds or, simply, interpretation. The latter is now acknowledged as a key mental faculty, as important as naive physics, naive biology, naive arithmetic, and language. Like these other faculties, interpretation appears to develop out of an innate core that is domain-specific, well scheduled developmentally, and most likely evolved as an adaptation to social life (Baron-Cohen 1995; Perner 1991; Wellman 1990; Whiten 1991; also Bogdan 1997). Importantly, basic interpretive skills have their own brain localization in the medial frontal cortex, suggesting a genetic basis with an evolutionary history. Psychological and neurological research indicate that rudiments of interpretation can be found in nonhuman primates (Frith and Frith 1999; Tomasello and Call 1997).

Less known but as important is the fact that interpretation is also a *mind builder*, in at least three senses. It appears to *stimulate* the development of other mental abilities, such as imagination and planning in the social domain (Whiten and Byrne 1988). It may *redevelop* some abilities by turning or exploiting them in new directions. For example, pretense may turn out to be an interpretational reutilization of imagination and play. Interpretation also appears to redevelop self regulation in the direction of self control and

metavolition (Perner 1998). But the most dramatic and far-reaching role of interpretation is *designing* new mental faculties. I construe this mind-design role in the sense that the tasks and often even the modus operandi of interpretation organize and constrain the tasks and the modus operandi of the faculties it brings about in development.

The mind-design role of interpretation revolves around and gradually builds on the generic sense of reference that enables the child to recognize intentionality or aboutness. This sense of reference develops in stages that help bring about some unique cognitive abilities, such as communication by shared meaning, the mastery of word meaning, the metarepresentation of mental states, and metamentation or thinking about thoughts, in this order. Each new ability in this developmental sequence stands and builds on the shoulders of its predecessors, which suggests a gradual construction of the human metamind (Bogdan 2000). So argues this paper.

Its argument combines three distinct but converging lines of analysis – conceptual, psychological and evolutionary – each necessary but not sufficient to carry the weight of the conclusion. The conceptual analysis reveals *what* the generic sense of reference is, in terms of the tasks involved and the categories and schemes that handle the tasks, and shows that such a sense of reference is at the heart of communication by shared meaning, word reference, metarepresentation and metamentation. Psychological data track the developmental parallels between interpretation and the abilities it gradually designs; these parallels reveal the constructive role of interpretation by marking when and sometimes how the design work was done. Finally, an evolutionary hypothesis looks at why interpretation managed to design new cognitive abilities around a sense of reference. I conclude with some remarks about how the argument of this paper can help an evolutionary explanation of the uniqueness of the human mind.

2. DESIGN BY TASK EMULATION

To ascertain that interpretation designed a mental ability, the conceptual analysis must show that the tasks of interpretation are constitutive of or structurally close or even isomorphic to the tasks of the ability in question. If either of these relations holds, I say that the latter tasks *emulated* the former. Task emulation is the conceptual core of the mind-design hypothesis. It shows that what a designed ability does is very much like, or crucially based on, what interpretation does. I develop the argument in the ontogenetic order in which interpretation is supposed to have designed communication by shared meaning, mastery of word reference, metarepresentation and metamentation around a growing sense of reference. The

description of each of these abilities will be minimal, just enough to make the case for design by interpretation; only communication by shared meaning receives more attention because it introduces most of the elements of the analysis. Little is known about how the task emulation is accomplished psychologically but I venture a few speculations in section 4.

2.1. *Communication by Shared Meaning*

Communication by Shared Meaning, or CSM in short, begins prelinguistically by way of actions, gestures, pointings, exchanges of looks or facial expressions, bodily postures and later spoken utterances. For simplicity, I group these means of communication under the notion of *acts*. CSM is heavily indebted to interpretation. Since the best known version of this debt is Grice's account of meaning, I will use it to frame our discussion. On Grice's account,

(a) *acts mean* something because communicators mean something; and
(b) *communicators mean* something by an act if and only if they intend the act to produce some *mental* effect – e.g., attention, emotion, belief – in an audience by means of the audience's *recognition* of this *intention* (to produce the effect in question)

Given the tumultuous debates around the Gricean account of meaning, some caveats are needed. Since I begin with prelinguistic communication, the more general notion of communicator replaces that of speaker. Furthermore, the interpretive categories cited in (a) and (b) need not be sophisticated. Prelinguistic infants do *not* recognize thoughts and beliefs, let alone intentions, yet they do interpret emotions and seeing, and they *do* communicate by drawing on the representations of such relations. Another caveat is that CSM need not require (as often feared) a regress of mutual recognition of intentions, nor therefore higher-order metarepresentations of propositional attitudes. Neither adult nor infant communication need be involved in such regress or recursion, unless normal strategies, conventions and literal uses are violated; only older children and adults recognize such violations and can do something about them (Perner 1988; Leekam 1991). All that is needed is a shared environment in which what is communicated is made manifest through exchanges that yield a mutual recognition of what is shared (Sperber and Wilson 1986; Gomez 1998).

With these caveats on record, I return to Grice's analysis. I take it to ground communicated meaning in interpreting others. I parse this idea as follows. To recognize that an act (action, gesture, utterance) means something is to recognize it as intentional. It is the recognition of *act-intentionality*. To recognize that one means something by

recognizing one's attitudes is to recognize *agent-intentionality*. On the Gricean account, the recognition of act-intentionality is based on that of agent-intentionality. There is nothing in this account that prevents communication from being *nonreferential* by merely conveying to an audience an attitude of the communicator. This is as it should be. Communication can be bilateral and topicless, through exchanges of attitudes. This is how human communication begins and why it develops as it does. It can be argued that Grice's analysis also fits the utilitarian and imperative communication of apes, when (say) voluntary grunts and gestures reveal some condition of the communicator (pain, pleasure) or anticipate some behavior and are recognized as such by the audience, without any referential import in the outside world. When communication becomes referential, apparently only in humans, it is because attitudes are recognized as being related to conditions of the world (Bruner 1983; Adamson 1996; Tomasello 1996). That is a development not captured by Grice's analysis. Let us see what this new development entails.

To recognize that an act *means* X, one must recognize the *relation* between the act and X, the *direction* of that relation and its *target*, i.e., X itself. The recognition amounts to representing act-intentionality by means of a *metaintentional* scheme that links up the categories involved – relatedness, direction, target. I parse the recognition of act-intentionality into these three categories because it makes conceptual sense (these *are* prerequisites of a sense of reference) and also because, empirically, it turns out that interpreters and hence communicators may possess some of the categories but not others, with corresponding limitations in their sense of reference. In short, there is *recognition* of an act-about-world relation when the act's relatedness, its direction and target are individually but also jointly recognized.

If a sense of reference builds on the recognition of act-intentionality and the latter draws on the recognition of agent-intentionality, we should expect the component categories of relatedness, direction and target, and hence the rudiments of a sense of reference, to originate in the recognition of agent-intentionality. This is where interpretation comes into the picture because it alone represents agent-intentionality as a subject-world relation. In so doing, it provides the categories that constitute a sense of reference. My hypothesis (developed in Bogdan 2000) is this. The recognition of *relatedness* (more exactly, a purposed or active relatedness) is likely to originate in the recognition of agency or goal-directedness. Many animal species recognize agency by recognizing that an organism actively relates to the world (is alive and about), as opposed to being inert, dead or otherwise unconnected. The recognition of agency may initially

belong to a naive biology before being appropriated by interpretation and integrated into its conceptual gadgetry. The recognition of the direction of agency appears to rest mainly on the ability to detect bodily orientation and to follow gaze. Few species, perhaps only the higher primates, track gaze. There is evidence of a brain-specialized mechanism for such tracking (Baron-Cohen 1995; Frith and Frith 1999). Although chimpanzees and other great apes are credited with the recognition of relatedness and its direction, through gaze following, there is skepticism about their ability to individuate targets (Povinelli 1996; Tomasello and Call 1997). They appear to identify the target of gaze egocentrically, in terms of their own perception and motivation, as opposed to exercising a specialized and context-invariant skill. This limitation may be connected with the inability to extract from gaze its attention value as a sign of internal or mental focus. This is why we can rate ape interpretation as *interactive* because it picks out only observable subject-world relations in terms of features of the environment and of conspecific behavior and assumes widely shared goals and standardized behaviors to reach them and also because it is employed essentially in an utilitarian manner (Tomasello and Call 1997; also Bogdan 1997; 2000). Let us take stock of what was said so far with the help of a diagram (Figure 1).

An interactive interpretation fails to deliver a recognition of target because, as noted later, it fails to secure a sharedness of attitudes and experiences. This is why ape communication does not develop to be meaning-sharing. It may appear wrong-headed to derive a sense of reference from sharedness of attitudes and experiences. One would have thought that the derivation goes the other way, from an ability to refer to something to sharing attitudes and experiences about that something. How can one share with others an attitude or experience about a target unless all participants already know how to refer to the target? What is communication if not mapping an intention to represent something into the communicative intention to have an audience recognize that intention to represent? (Searle 1983, chapter 6). The issue, however, is not representing something (intrinsic intentionality) versus sharing attitudes, experiences and meaning. As noted already, most animal minds are intrinsically intentional, hence represent targets and may even share information about the world. The issue is whether those same animal minds also *recognize* (not just have) intentionality and thus the meaning relation itself. The notion of intention to represent (Searle's) is ambiguous between these two readings. If the intention amounts to representing something, then it is a case of intrinsic intentionality. But if the intention presupposes *recognition* of intentionality and thus a sense of reference (as I suspect it does), then it

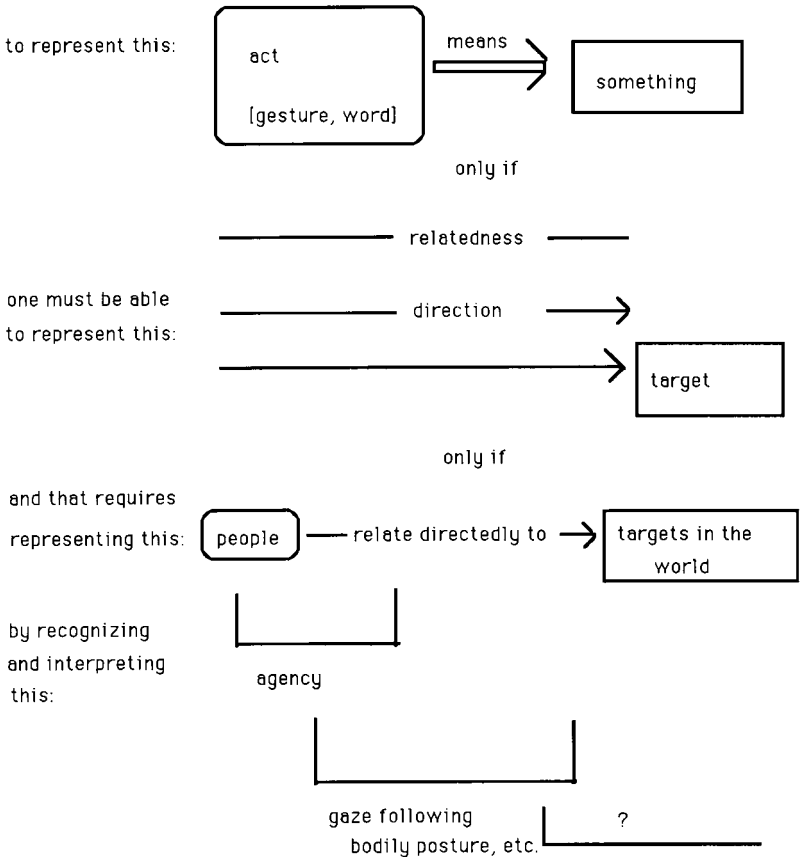

Figure 1. Act-meaning by people-meaning.

would be wrong to think of it as prior to and independent of interpretation and communication. Conceptually, a sense of reference depends on interpretation, and, developmentally, it grows out of a sense of shared experiences.

The analysis, diagrammed in Figure 1, suggested that the sense of reference provided by the interactive interpretation of apes is incomplete because it lacks the category of target. The proposal now is that the category of target emerges gradually from sharing mental stances and experiences. As follows. The key to this development is a new, *inter-subjective* form of interpretation, apparently uniquely human. It operates by exchanging emotions, experiences and attitudes, and by using such exchanges to inform others about the world and themselves. Infants are known to bond sentimentally with adults, mothers in particular, by exchanging emotions, looks, smiles, and other expressions of inner states. The exchanges are first bilateral and topicless, aimed at securing an *interper-*

sonal relatedness that expresses psychologically the deeper physiological phenomenon of mother-infant regulation (Trevarthen 1993). Around nine months, infants begin to engage in *social referencing*, whereby they use such bilateral exchanges and gaze following to triangulate proximal targets, such as an action by the infant or some nearby event. Around eighteen months infants begin to *share attention* with others about distal targets, such as specific objects or properties (Bruner 1983; Hobson 1993; Trevarthen 1993; Adamson 1996; Tomasello 1996). Such intersubjective interaction and communication take place prior to and during the earliest phases of language acquisition. These advances in interpretation and CSM are systematically linked to a growing sense of reference, in the sense that the targets the infant triangulates with the help of others become more distal and better individuated. The semantics of the first words follows the same pattern and developmental schedule (Baldwin 1991; Adamson 1996; Tomasello 1996). This entire process draws on the infant's ability to interpret someone's stances and experiences as takes or comments on something of mutual interest, a sort of topic-comment strategy or *topical predication*. It is an extraordinary ability, quite unique and with far-reaching consequences for mental development (Bogdan 2000, chapter 3). The initial bilateral and topicless exchanges accustom the infant to the reactions of the others and allow a coordination of reactions qua potential takes or comments. When proximal and later distal topics are introduced, a framework is already in place that allows the infant to recognize what another person relates to in response to what the infant relates to (comments on) and vice versa (Bruner 1983; Tomasello 1996).

This, roughly, is the developmental matrix in which the infant acquires a sense of shared meaning. In both social referencing and shared attention the infant's looking or gesturing and smiling or frowning at mother and then at some action or object and deliberately causing mother to look in the same direction and smile (or frown) back in recognition is a case of acts-meaning-something (e.g., 'what-about-my-action?' or 'isn't that object fun?') because the infant so intends her acts and mother so recognizes them. The meaning is what is shared, not just as a mere action or object but as what a mental take or evaluation is directed at. The bilateral sharing is achieved by producing and recognizing emotion- and experience-driven takes on the world and on each other. What counts as act-meaning (e.g., what a look or smile means) at these early ontogenetic stages is determined by agent-meanings. These meanings are negotiated through interpretation in exchanges of mental stances (comments) on topics of mutual interest. In other words, what is meant is what is shared, and what is shared is what is interpreted intersubjectively in exchanges of mental stances through top-

ical predications. The earliest generic sense of reference is built into this matrix of shared meaning. The potential referents are items in the topic slot awaiting language and conceptual developments for sharper individuation.

In sum, a generic sense of reference builds first on a recognition of agent- and then of act-intentionality. Those recognitions are effected by means of three interpretive categories – relatedness, direction and target. Apes and some autistic children have a limited sense of the former two categories and lack the third, in its wider format of topic, by failing to share experiences and attitudes as comments. As a result of this failure of intersubjectivity, they cannot communicate by shared meaning or not communicate well. Intersubjective children master the category of topic and hence the format for recognizing the target of intentionality. Thus, task by task and category by category, differences in interpretation translate into differences in recognizing basic intentionality and acquiring a sense of reference and therefore into differences in communication by shared meaning (Figure 2).

Details aside, so far the development of a sense of reference went from subject-world interpretation to prelinguistic communication by shared meaning. The next stage is building the semantics of language on this basis.

2.2. *Symbolization and Word Reference*

The acquisition of symbol and word semantics exploits naturally the communication by shared meaning and the sense of shared reference it affords. So construed, the language-acquisition process is less abrupt, less radically new, and less surprising than it may seem on other accounts. Once children and adults engage in social referencing the opportunity arises to use new and often arbitrary acts, such as mutually recognized gestures or sound patterns, to *symbolize* shared meanings. This becomes symbolic communication by shared meaning. Symbols are *first* introduced prelinguistically in interpretationally formated contexts of social referencing and later shared attention. In those contexts symbolization is likely to be first construed by infants in terms of adult *comments* on topics of mutual interest – a game that infants already know how to play (Bates 1976; Bruner 1983; Hobson 1994; Tomasello 1996). In Gricean terms, then, gestures and words are acts that symbolize something because agents intend them to, and agents intend them by commenting in this way on shared topics. Act-symbolization thus builds on agent-symbolization and the latter in turn is viewed by the infant as a new version of agent-intentionality, with new stances and attitudes to what is shared and communicated. The conceptual parallel between symbolization and interpretation-based topical

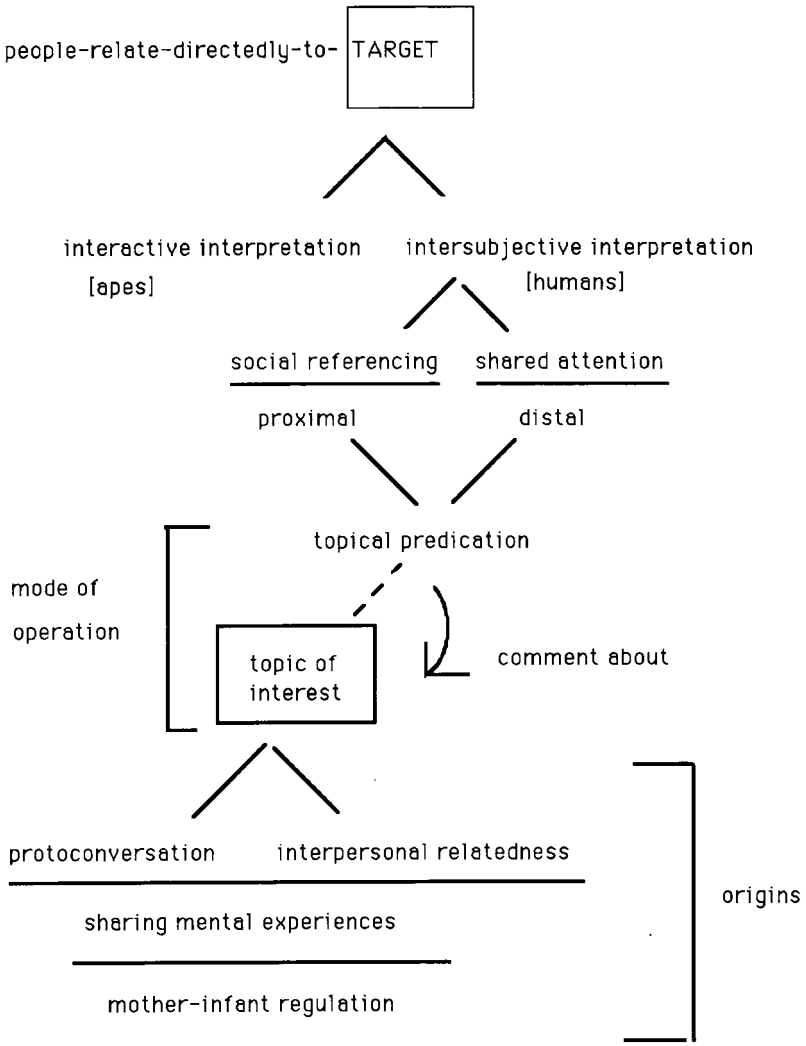

Figure 2. Target fixation by intersubjective interpretation.

predication is diagrammed in Figure 3 (taken from Akhtar and Tomasello 1998).

This line of analysis establishes that the initial tasks of symbolization (gestural or spoken) are based on those of communication by shared meaning and therefore emulate the tasks of interpretation. This emulation claim can be parsed as follows: one could not learn what words and other symbols mean unless one were an intersubjective interpreter of other people and had a sense of their intentionality and in particular of their referential intentions; and one could not master word and symbol reference unless

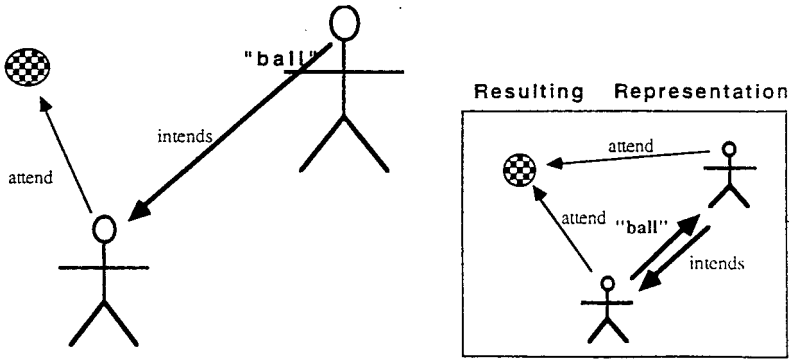

Figure 3. The symbolization game.

one could also handle topical predication. These are crucial (though not the only) prerequisites for mastering the semantics of a language. The appeal to emulation is not meant to explain how this mastery process actually works. Other abilities must bring their contribution (Bloom 1997). Thus, grammar helps categorize words along various dimensions while the child's naive physics or biology help recognize the kinds of items in the topic position which can be individuated separately and thus given distinct names. These other abilities further individuate for the child the specific targets of the adult's referential intentions and naming practices. The point is that the initial sense that symbols and words refer because people do can originate only in the child's ability to interpret other people, their original intentionality and the derived intentionality of their acts.

2.3. *Metarepresentation*

I noted that the semantics of language could not be normally acquired without subjective interpretation and the resulting communication by shared meaning. As they get assembled in early and middle childhood, the semantics and other dimensions of language in turn enable interpretation and communication by shared meaning to acquire new depth and complexity. A crucial outcome of this development is that interpretation and the sense of reference it sponsors reach further into the mind, beyond the perceptual world of the here-and-now. Around 3 to 4, children are able to represent false belief, a milestone in interpretation because it is the representation of a mental state that is no longer controlled exclusively by the environment. For the first time the child becomes aware of what is in the mind of another person, irrespective of how the world is and is perceived to be. Such a representation of mind-world relations as propos-

itional attitudes begins with simple desires and beliefs and is gradually extended to intentions, thoughts, and complex attitudes derived from them, such as hope, regret, shame, and so on (Astington and Gopnik 1988; Perner 1991; Bogdan 1997). The child's sense of reference moves from perceived acts to invisible mental states. The metarepresentation of attitudes now provides a sense of mental reference. The *mental* components of attitudes are understood as referential devices and later as self-referential as well.

2.4. *Metamentation*

Finally, reflexive thinking or metamentation is another uniquely human ability designed by interpretation. The longer argument, book-length, is developed elsewhere (Bogdan 2000). As with the previous abilities, my aim here is merely to outline and illustrate the emulation thesis. So here it goes.

Metamentation is the ability to form mental representations about other such representations and in particular thoughts about one's thoughts. The latter version is metathinking. It amounts to thinking explicitly about thoughts *as* mental structures that represent something. *Simple* metathinking is representing a thought as a mental relation to a fact, as in representing my thought [that Vienna could have been a Turkish city] as false but not that implausible. *Reflective* metathinking is specifying a thought in terms of other thoughts, as for example when thinking that [a + b = b + a] by thinking that [addition is commutative] and that [the first thought represents a case of addition]. There are other forms of metamentation, such as imagining situations populated by people who think about other people's thoughts, also metamental predication, which is topical predication in which the topic is one thought and the comment another, and recursive metathinking, in which thoughts are embedded in other thoughts, embedded in still other thoughts, and so on. Since all these forms of metamentation share the core ability to think about one's thoughts, metathinking would suffice here to illustrate the emulation argument.

Metamentation relies on both the recognition of intentionality and the intersubjective topical predication, hence on shared-attention tasks, as much as communication by shared meaning and the acquisition of word reference do. So the earlier analysis applies here as well and thus establishes part of the emulation argument. The reason is simple: one could not represent a thought as a mental relation to some content (factual or mental) unless one represents that relation as intentional, hence in terms of relatedness, direction, and target. A good deal of metamentation involves evaluation, deliberation or planning, and none of these would work unless one were also able to comment explicitly with a thought about another

thought as topic. Developmentally, we saw, one could not master topical predication without being able to enter into an intersubjective contact or sharing of mental experiences with another person.

This diagnosis takes us as far as shared-attention tasks, which are a distant basis for metamentation and many other abilities besides. Metamentation also relies on further interpretive abilities that mature later in childhood. Thus, one cannot metamentate unless one can also do at least three things: *metarepresent*, in the sense that one understands that a mental representation can be true or false of a target, can vary in how it represents a target from person to person or by the same person at different times, and so on; recognize and track *iterated* embeddings of mental representations about other mental representations; and entertain *explicit metathoughts*, which are thoughts explicitly recognized as mental structures that represent and whose content is specified relationally in terms of other thoughts. It is easy to see now that all these further conditions on metamentation cannot be met unless one has a grasp of mind-world and mind-mind relations and hence a sense of mental reference that only interpretation can provide. These points are summarized in the next diagram (Figure 4).

Let us take brief stock. Four key cognitive abilities – communication by shared meaning, mastery of symbolization and word reference, metarepresentation and metamentation – were shown to branch out of a generic sense of reference that originates in interpreting first other people (agent-intentionality) and later their acts (act-intentionality) and minds (mind-intentionality or metarepresentation). The abilities in question are new and owe their existence to interpretation, even though they require other contributions in order to operate. This is the extent to which interpretation is a mind designer and in particular a designer of abilities that revolve around a generic sense of reference.

Is there psychological evidence to support this hypothesis, so far defended mostly conceptually? I think so and introduce it in the next section.

3. PSYCHOLOGICAL EVIDENCE

I divide the evidence into (a) systematic *correlations* between key acquisitions and changes in interpretation *and* significant novelties in the abilities suspected to emulate interpretation; and (b) symptomatic *failures* (of the dog-that-didn't-bark sort; the only sort of dogs I like), where the absence of or some deficit in an interpretive ability explain the absence or deficit in a new ability suspected to emulate interpretation. I propose to sample

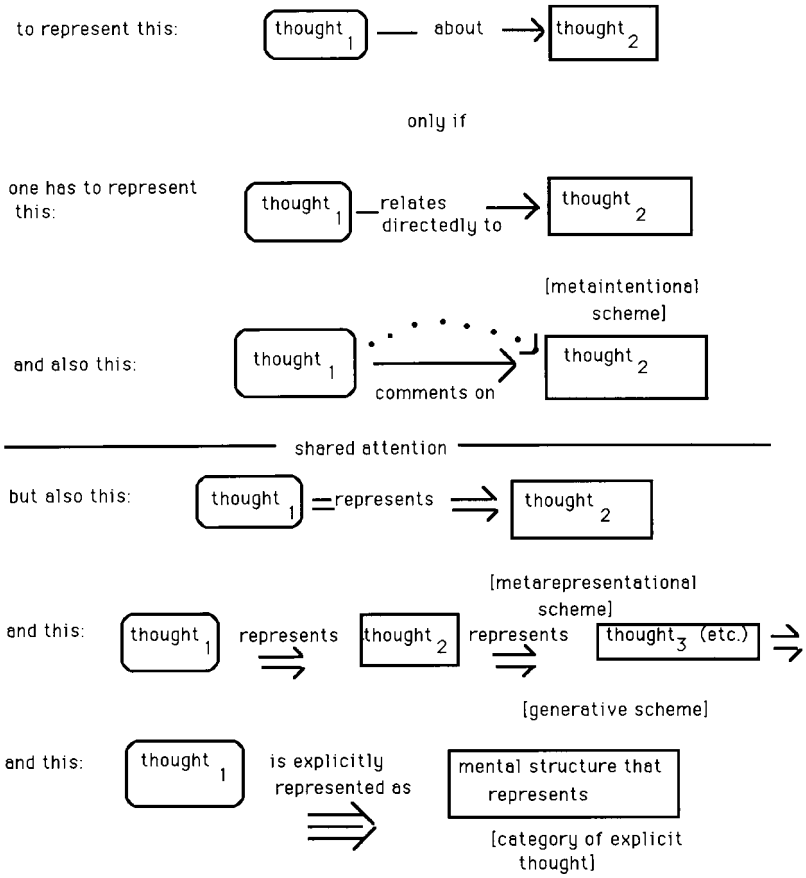

Figure 4. Metamentation.

such correlations and failures by returning to the conceptual analyses of the previous section and putting them in historical motion.

Consider first the analysis diagrammed in Figures 1 and 2. It was noted then that apes recognize agency-based relatedness and its direction but not its target, a fact suggesting rudiments of an incomplete metaintentional scheme and hence an incomplete sense of reference. As a result, it is doubtful that apes can share and communicate about specific and variable meanings, aside from widely shared goals, such as food, dangers, mating, and the like, probably built into their sensorimotor cognition. For roughly the same reasons, it is doubtful whether apes can truly master fragments of symbolic languages. Over the years a few captive apes growing up in a human culture have been trained or learned by themselves to communicate in novel ways, often through figurative symbols. Kanzi is a recent example (Savage-Rumbaugh and Lewin 1994). This is not the place to develop a

diagnosis of this remarkable phenomenon. With an eye to the theme of this paper, I will say this much. Natural abilities of animals are displayed in the wild. And there apes do not recognize intentionality fully and do not communicate by shared meaning, let alone symbolically. In captivity some enculturated apes may recruit existing capacities to handle new challenges. What needs to be shown is that captive apes develop a full sense of reference, instead of using their partial sense of agent-reference and other natural abilities to solve the new problems posed by symbols expected to link up with new behaviors.

I ascribed the limitations of ape interpretation to its interactive character and its failure to interpret intersubjectively and thus to relate interpersonally and share referencing, attention and meaning. Possibly a matter of degree, this failure seems sufficient to hamper the acquisition of a full sense of reference. It turns out that a somewhat similar failure might explain the limits of autistic interpretation, with serious consequences for their communication by shared meaning and word semantics. Both apes and autistic children express emotions and other experiences in communication and may intend some acts to catch the attention of others and direct it to some item of interest, thus displaying some meaning-directed communication. What seems to be missing, though, is the sharedness, mutuality and turn-taking present in the interpersonal exchanges and communication of normal human children with others. One important casualty is topical predication. Its absence prevents apes and many autistic children from identifying items of shared interest and as a result affects their ability to master word reference. Autistic children also have difficulties with propositional attitude and metarepresentation in general, which limits their sense of mental reference and the ability to metamentate (see Baron-Cohen et al. 1993 for a survey).

Looking now at normal child development, one can also detect a number of supportive correlations. Consider first the acquisition of word semantics. As noted earlier, many factors must be involved in this process, including advances in grammar, the perception of salient perceptual attributes and of functional affordances and the development of the naive theories of various domains (physical, biological, numerical) whose denizens get named. Yet the whole process would remain mysterious if it could not build on a prior, prelinguistic sense of reference shared by infants and adults in their communication (Bruner 1983; Tomasello 1996; Bloom 1997). This growing sense of reference must have something to do with the fact that the earliest words are generally acquired around the first birthday, when social referencing is in place, with a tentative and imprecise reference that matches the vagueness and proximality of social referencing.

The acquisition process speeds up and delivers a sharper reference around 18 months, when shared attention takes over and delivers a more precise, focused and distal reference (Baldwin 1991).

Consider metamentation next. My analysis predicts that it builds on a succession of interpretive (and other enabling) skills and therefore should not be available to children before this construction is in place (Bogdan 2000). For example, thinking about thoughts requires a full sense of metarepresentation, courtesy of the categories of propositional attitudes, and therefore cannot develop before 4; and it doesn't. Metamentation also requires a generative scheme that allows thoughts to be embedded in other thoughts. This generative ability matures rather late, after 7 or 8, following closely the maturation of the generative metarepresentation of propositional attitudes (Perner 1988; also Bogdan 2000).

Another relevant piece of evidence is the delay between interpreting others and oneself. Unlike the intentionality of one's agency, the intentionality of one's cognition is not transparent to the child right away. A relational understanding of the intentionality of mental acts takes time to develop. Such an understanding appears to begin with others, not with self. From gaze to seeing, believing or remembering, the relations or attitudes of others are interpreted before those of self (Astington and Gopnik 1988). As Perner puts it, "children do not come to understand the relevant aspects of their own mind any earlier than they understand the relevant aspects of other people's minds" (Perner 1991, 270). It is by interpreting other minds that children first acquire a sense of reference and that is prior to acquiring a sense of self-reference.

4. REASONS FOR MIND DESIGN

Suppose that, rough and programmatic as it is, the story told so far has conceptual and psychological plausibility: a generic sense of reference is gradually formed by interpreting other minds; once installed in the young mind, that growing sense of reference helps interpretation design novel mental faculties, such as communication by shared meaning, mastering word reference, metarepresentation and metamentation, in that developmental order. This story invites the next and further-probing question: Why is interpretation a mind designer and why does it design the faculties just cited around a generic sense of reference? I do not have a full answer to this question but I have a hunch that the following line of inquiry is worth pursuing in the search for an answer (Bogdan 2000).

To see where the hunch is coming from, I propose to step back in evolution and ask another question. Among all forms of primate cognition, *two*

are widely thought to have spawned pressures for higher mental faculties. These are the forms of cognition involved in tool making and use (in the technological domain) and conspecific interactions (in the social domain). Why would technological and social cognition be most likely to shape the primate mind? Because, I think, they have (at least) four interesting properties that make a unique and very potent combination:

(a) they are dedicated to *instrumental intervention* in their domains (i.e., using an object or conspecific as instrument or means to reach one's goals);
(b) instrumental intervention is of the *cause-causation* sort (i.e., causing an instrument to cause further effects that attain one's goals);
(c) the instrumental intervention of the cause-causation sort becomes more successful and efficacious (hence selected for) when the ways and means of intervention are *mentally anticipated*, *controlled*, *manipulated* – in short, *rehearsed* – before an action is initiated (as in Figure 5).

As a result,

(d) the ways and means of intervention – call them *instrumental representations* – are themselves re-represented in some form, which is to say that the instrumental representations become *targets* of mental activity and hence of it metainstrumental cognition (as in Figure 6).

In an earlier work (Bogdan 1997) I argued that interpretation *is* a form of instrumental cognition in the social domain and that it satisfies conditions (a) and (b). In a later work (Bogdan 2000) I further argued that when interpretation also meets conditions (c) and (d), there emerges a powerful *mind-design* formula. I first sample some of its basic parameters before bringing interpretation back into the story.

Think of an *instrumental domain* of cognition as populated by patterns of relations linking an agent's actions to instruments (physical or social), states of the world and outcomes as goals. In a mental-rehearsal mode, metainstrumental cognition allows an agent to entertain instrumental representations of various, often novel and possibly counterfactual patterns of action-instrument-world-outcome relations. Each slot (agent, act, instrument, world, outcome) in this scheme is potentially a *variable* in the sense that its identity, perhaps internal complexity and relation to other slots can vary and change according to context, interest and learning. As a result, the mentally rehearsed instrumental representations that anticipate actions of the cause-causation sort can be changed, linked to other representations, brought under new categories and schemes, and perhaps made explicit – in short, *re-represented*. With other conditions are in place, such re-

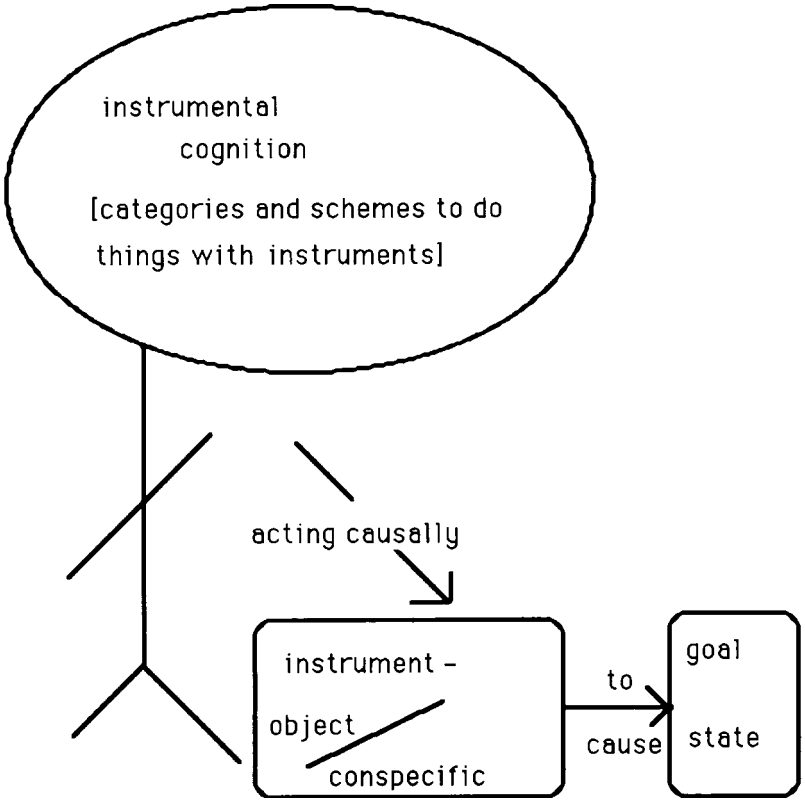

Figure 5. Instrumental cognition.

representations may create a blueprint or model for the evolution of new cognitive abilities. Roughly, as follows.

Consider the social domain. An interpreter A can influence or manipulate a subject B by acting causally upon B's intentional relations to the world. Intentional relations *are* the instruments causally manipulated in the social domain, that is, caused to cause situations and events that meet the interpreter's goals. Thus, for example, in order to influence a subject's attention in a desired direction, an interpreter may turn and look intently or point in that direction. The strategy works because the interpreter categorizes attention as a subject-world relation and can re-represent the aspects that reveal attention as targets of interest and causal intervention when rehearsing some plan of action. In general, in its earliest and simplest forms, interpretation represents and rehearses agent-world relations or agent-intentionality, as manifested in sundry behaviors or visible relations to states of the world, in either interactive or intersubjective forms (as illustrated in Figures 7 and 8).

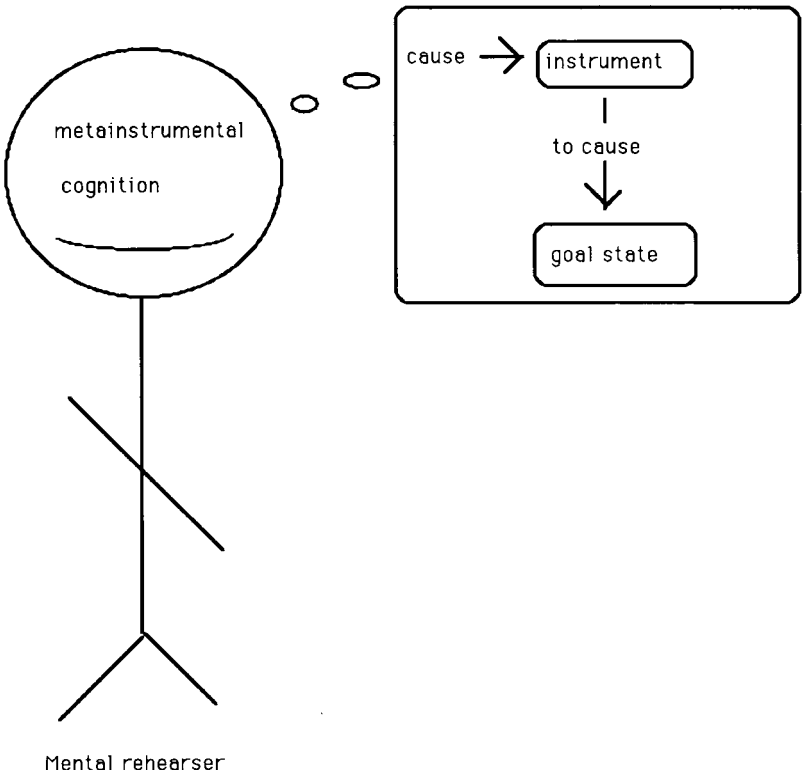

Figure 6. Metainstrumental cognition.

The next critical step in cognitive development is made when specialized *acts*, such as gestures, pointings or vocal signals, become regularly associated with observable patterns of agent-intentionality and the situations in which the patterns are manifested. At that point agent-intentionality can be reinterpreted and targeted for intervention of the cause-causation sort *as* act-intentionality. That can be done not only because the representation of agent-intentionality brings about that of act-intentionality, as argued earlier, but also because the interpreter is intent on causing others to cause situations she desires and the act-world relations – or, more elaborately said, the relations of act-meaning-something-in-the-world – have become new levers of causal intervention. Act-world relations have to be used causally to get what one wants. Captive and enculturated apes and equally captive and enculturated human infants use the acts (gestures, symbols, utterances) to affect causally the intentionality of adults and get them to do what is desired.

The representation of the act's meaning thus develops out of the representation of agent-intentionality because of the practical necessity of causal

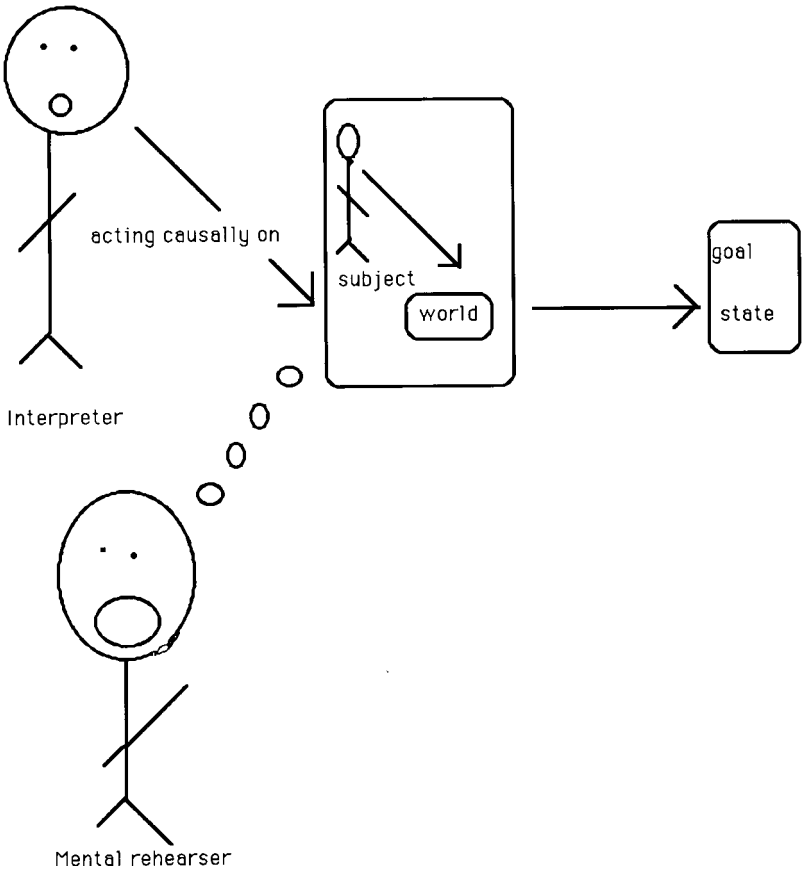

Figure 7. Interactive interpretation, with mental rehearsal.

intervention. The meanings of acts get represented because they can be caused to cause desired effects. As long as the ape and child versions of interpretation remain situated, the representation of act-intentionality is likely to be implicit, indirect and perceptually interwoven with that of agent-intentionality, just as the latter is interwoven with the perception of the situations and behaviors that reveal it. The figurative symbols that some captive apes learn to recognize and employ are probably represented in this mixed act-agent-world way and in terms of the specific classes of behaviors and situations that the symbols anticipate. The same may be true of how human infants first represent gestures and sound patterns, soon to be understood as words. To the extent that apes or young children can also re-represent such acts-world relations in mental rehearsal, they may reenact, perhaps imagistically, some aspects of their meanings and anticipate causal interventions on such aspects. A partial sense of an act's

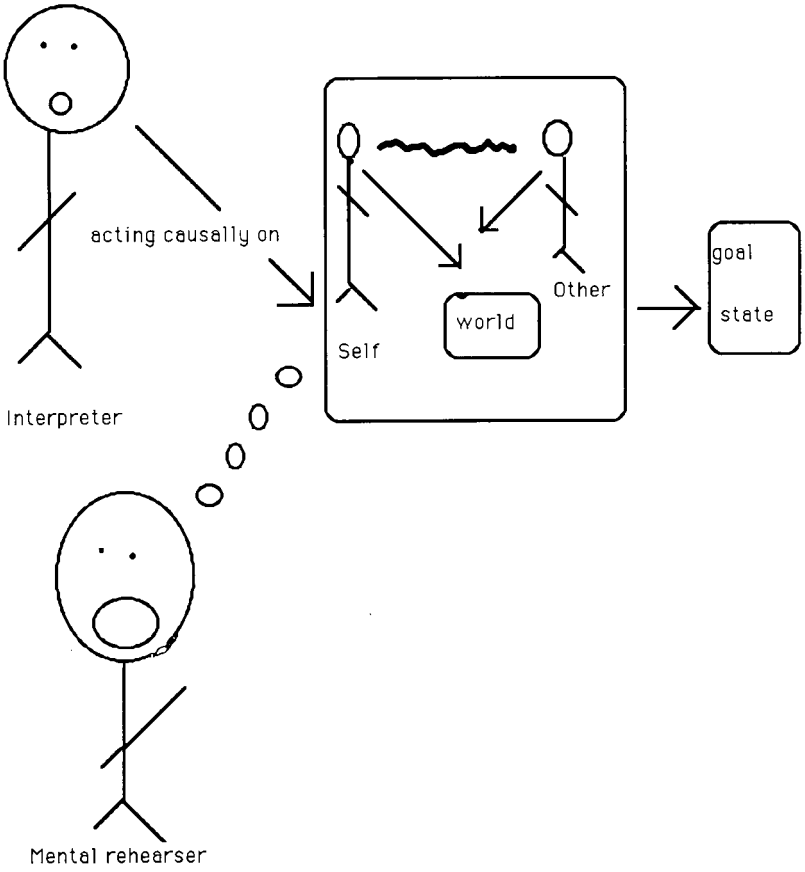

Figure 8. Intersubjective interpretation, with mental rehearsal.

reference thus becomes available to mental calculation, in imagination and pretense. The result is a purposeful and planful communication that, in the intersubjective minds of human children, allows the completion of the metaintentional scheme as act-reference and thus helps considerably in the acquisition of a semantics for the language.

An important point is worth making at this juncture. In the early stages of language acquisition, the child *also* begins to treat the re-representations of act-intentionality as higher-level internal stand-ins or simplifiers of more complex lower-level representations of action schemes that, in the social domain, track patterns of agent-intentionality. These internal stand-ins may be called *symbols from inside* – somewhat like icons on a computer screen which stand for, and activate when clicked, complex lower-level procedures. The phenomenon of cognitive simplification or clustering by internal symbolization is well known to evolutionary theorists and com-

parative psychologists (Bogdan 1997, 142–148), squares with the classical cognitive-scientific view about what counts as a symbol in computational processes, and also squares with the Piagetian notion that some internal aspect is a symbol if processed by a higher-level scheme to activate lower-level action schemes in the absence of real-world objects (Bates 1976).

Many theorists are tempted to think that symbolization from inside suffices to explain language acquisition. I doubt it. If the general argument of this paper is right, the grasp of symbolization and word semantics also requires a sense of reference, acquired only by interpreting other minds. And if the specific argument of this section is right, the grasp of word semantics emerges precisely because, in goal scripts and mental rehearsals dedicated to causal intervention in the social domain, the interpretation-based sense of reference meshes with symbolization from inside, as complex patterns of agent-intentionality are simplified and re-represented by internal symbols. Whatever underlies the understanding of words, it must at least do double duty: (i) activate concepts, memories, images, and other mental structures and procedures through symbolization from inside and (ii) align the result to a generic sense of reference. If apes and other well trained animals seem to understand and act on figurative symbols and words, it is most likely to be because the training has supplied them with some form of symbolization from inside by way of images or sounds they associate with definite classes of behaviors and situations. It is unlikely that these prehuman symbolizers connect that achievement with a genuine sense of reference.

This was brief and sketchy but I hope sufficiently suggestive of the main idea: the primate minds acquire a partial or complete sense of reference because it causes subject-world relations and later act-world relations to cause desired effects; as a result, the primate minds evolve categories and schemes to represent such relations. The active causal involvement (of the cause-causation sort) explains the development of the mental equipment – in the form of categories, schemes and procedures – required to do the job. Mental rehearsal with the representations generated by such categories and schemes in turn creates the opportunities for novel re-representations and that opens the way to the emergence of new cognitive abilities based on a sense of reference. This is how I would explain the transition by task emulation from interpreting others to communicating by shared meaning to mastering the semantics of a language to metarepresentation to metamentation.

5. REDUCING THE EXPLANATORY GAP

The most perplexing question in evolutionary psychology is how to explain the entirely novel cognitive abilities of humans, such as a grammar-based and semantically fine-grained language, imagination, creative thinking, self reflection and self consciousness. These abilities seem absent in non-human primate or any other species and to that extent do not seem to have evolved, certainly not by an incremental and time-consuming natural selection spanning many species. This reasonable estimate led many cognitive scientists to conclude that such unprecedented abilities are either the outcome of a unique evolutionary accident or an emergent neural property of a complex brain or the result of an equally unique capacity to learn new things. Neither conclusion is very convincing, for various reasons widely discussed in the evolutionary literature (Barkow, Cosmides and Tooby 1992; Dennett 1995). One powerful reason is that these explanations bear on abilities that show very complex design – usually, a hallmark of adaptation. Yet the puzzle remains: there is a huge and apparently unfathomable gap between what looks adaptively complex and well designed, on the one hand, and a total and apparently evolution-free novelty, on the other hand. How could an evolutionary explanation bridge this gap in a plausible manner?

I do not have an answer and this is not the place to even begin to articulate it. But I want to conclude with some forward-looking suggestions. Elsewhere (Bogdan 2000) I argued for the need to shift the frame and time of evolution from primate phylogeny to human ontogeny and from the selective pressures of nature to those of culture. What is unique about the human mind is its development and the sociocultural ambiance in which that development takes place. The unique abilities of the human mind owe a great deal to its development through culture. Since first and forcefully articulated by Vygotsky in the early decades of the 20th century, this position has been further developed (e.g., Bruner 1983; Tomasello 1996) and is gaining ground. The fact to note here is that both child development and culture *are* distinguished and respectable evolutionary phenomena (no unique, accidental, pure-emergence hocus pocus here), with some primate and animal pedigree. Development is a form and time-slice of evolution in all species and moves in new evolutionary directions in primates and humans. Culture is as real as nature and, like nature, can spawn strong pressures that shape human abilities, from eating and writing to imagining and thinking.

In the midst of this reframing and retiming of evolution stands the competence for interpreting other minds – perhaps the most important

cognitive ability the social primates evolve and certainly the most consequential for their mental development. This is where the argument of this paper reenters the picture to help reduce the explanatory gap between evolutionary history and the uniqueness of the human mind. For what the argument suggests is that several uniquely human abilities – such as communication by shared meaning, metarepresentation, and metamentation – developed first *as* forms or by-products of interpretation. And interpretation *is* an adaptation among primates, first evolved by natural selection and later, mostly in human development, by cultural pressures and inducements. According to the argument, humans communicate by symbols and language, metarepresent, metamentate, are conscious of their selves and their mental lives, and think creatively because, to a very considerable and decisive extent, they interpret other minds intersubjectively (Bogdan 2000).

The uniqueness of human minds builds on the unique shoulders of their intersubjective interpretation. But the latter is not a sudden, accidental, emergent evolutionary phenomenon, nor is it an open-ended, general-purpose learning faculty. Interpretation is a specialized adaptation with a distinct domain and a distinct mode of representation of domain-specific patterns of subject-world relations and, in its human version, a representation of mind-world, mind-world-mind, and mind-mind relations. It is these patterns of relations that actually structure the domains of the novel and unique faculties that intersubjective interpretation designs or helps bring about, as this paper has endeavored to show. The conclusion one should draw, I think, is that the human mind is unique in many ways not because it stands outside evolution or at its periphery but because, fully and centrally inside evolution, it piggybacked on an adapted competence with primate precedents that took a unique turn in child development, mostly under the watch of culture. Interpretation, development and culture are the main (but not the only) bridges across the explanatory gap between evolution and the human mind.[1]

NOTE

[1] I would like to thank several audiences of psychologists and philosophers who heard and reacted to earlier versions of this paper read at the University of Salzburg (special thanks to Josef Perner), University of Innsbruck (special thanks to Josef Quitterer), University of Arizona (special thanks to Keith Lehrer), University of Bucharest (special thanks to Mircea Flonta and Sorin Vieru), and my own Tulane University.

REFERENCES

Adamson, L. B.: 1996, *Communication Development During Infancy*, Westview, Boulder, CO.

Akhtar, N. and M. Tomasello: 1998, 'Intersubjectivity in Early Language Learning and Use', in S. Braten (ed.), *Intersubjectivity and Emotional Communication*, Cambridge University Press, Cambridge.

Astington, J. W. and A. Gopnik: 1988, 'Knowing You're Changed Your Mind: Children's Understanding of Representational Change', in J. W. Astington et al. (eds), *Development Theories of Mind*, Cambridge University Press, Cambridge.

Baldwin, D.: 1991, 'Infant Contribution to the Achievement of Joint Reference', *Child Development* **62**, 875–890.

Barkow, J., L. Cosmides, and J. Tobby (eds), *The Adapted Mind*, Oxford University Press, New York.

Baron-Cohen, S.: 1995, *Mindblindness*, MIT Press, Cambridge, MA.

Baron-Cohen, S., S. Targer-Flusberg, and D. Cohen (eds): 1993, *Understanding Other Minds*, Oxford University Press, Oxford.

Bates, E.: 1976, *Language and Context*, Academic Press, New York.

Bloom, P.: 1997, 'Intentionality and Word Learning', *Trends in Cognitive Science* **1**, 9–12.

Bogdan, R. J.: 1997, *Interpreting Minds*, MIT Press, Cambridge, MA.

Bogdan, R. J.: 2000, *Minding Minds*, MIT Press, Cambridge, MA.

Bruner, J.: 1983, *Child's Talk*, Norton, New York.

Dennett, D.: 1995, *Darwin's Dangerous Idea*, Simon and Schuster, New York.

Frith, C. D. and U. Frith.: 1999, 'Interacting Minds: A Biological Basis', *Science* **286**, 1692–1695.

Gomez, J. C.: 1998, 'Some Thoughts about the Evolution of LADS', in P. Caruthers and J. Boucher (eds), *Language and Thought: Interdisciplinary Themes*, Cambridge University Press, Cambridge.

Grice, H. P.: 1957, 'Meaning', *Philosophical Review* **66**, 377–388.

Hobson, R. P.: 1993, *Autism and the Development of Mind*, Erlbaum, Hillsdale, NJ.

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

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

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

Perner, J.: 1998, 'The Meta-Intentional Nature of Executive Functions and Theory of Mind', in P. Carruthers and J. Boucher (eds), *Language and Thought: Interdisciplinary Themes*, Cambridge University Press, Cambridge.

Povinelli, D. J.: 1996, 'Chimpanzee Theory of Mind?', in P. Carruthers and P. K. Smith (eds), *Theories of Theories of Mind*, Cambridge University Press, Cambridge.

Savage-Rumbaugh, S. and R. Lewin: 1994, *Kanzi*, Wiley, New York.

Searle, J.: 1983, *Intentionality*, Cambridge University Press, Cambridge.

Sperber, D. and D. Wilson: 1986, *Relevance*, Harvard University Press, Cambridge, MA.

Tomasello, M.: 1996, 'The Cultural Roots of Language', in B. M. Velichovsky and D. M. Rumbaugh (eds), *Communicating Meaning*, Erlbaum, Mahwah, NJ.

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

Trevarthen, C.: 1993, 'The Self Born in Intersubjectivity', in U. Neisser (ed.), *The Perceived Self*, Cambridge University Press, Cambridge.

Wellman, H.: 1990, *The Child' s Theory of Mind*, The MIT Press, Cambridge, MA.
Werner, H. and B. Kaplan: 1963, *Symbol Formation*, Wiley, New York.
Whiten, A. (ed.): 1991, *Natural Theories of Mind*, Blackwell, Oxford.
Whiten, A. and R. W. Byrne: 1998, 'The manipulation of Attention in Primate Tactical Deception', in R. W. Byrne and A. Whiten (eds), *Machiavellian Intelligence*, Oxford University Press, Oxford.

Department of Philosophy
Tulane University
New Orleans, LA 70118
U.S.A.