

Abstract

This chapter presents an evolutionary and developmental model, according to which intersubjectivity is intimately tied to *bodily mimesis* – the use of the body for communicative and representational purposes – to an extent that it can be said to co-evolve with it. I review some relevant primate evidence which shows that feral and captive apes are capable of the first two levels (involving e.g. empathy, shared attention and imitation), but not of the third level which involves an understanding of communicative signs, i.e. triadic mimesis. Enculturated language-trained apes, on the other hand, show some aspects of triadic mimesis, suggesting how our predecessors could have bootstrapped themselves to this level without language (and without a “theory of mind”). The emergence of language, on the other hand, opens the way to the highest two levels of intersubjectivity, bringing forth the understanding of “beliefs” and the use of folk psychology.

1. Introduction

Many if not most would agree that there is a close relationship between intersubjectivity and language. But what more precisely is this relation? The first impediment to answering this question is definitional. As several of the contributions to this volume show, there are rather different understandings of the concept of intersubjectivity. For the purpose of this chapter, intersubjectivity will be taken to be *the sharing of affective, perceptual and reflective experiences between two or more subjects*. As Brinck (this volume) notes, “sharing” takes different forms, some more immediate, while others more mediated by higher cognitive processes, e.g. what Barresi and Moore (this volume) call “understanding” as opposed to simply “sharing”. The phenomenon of *joint attention* (e.g. Moore and Dunham 1995), for example, would qualify as a paradigmatic form of perceptual intersubjectivity.

If it is difficult to reach a consensus concerning the notion of intersubjectivity, it is even more so when it comes to language! Therefore I will not here argue for but simply assume that language is *a conventional (normative) symbolic system for communication and thought*, where symbols are understood as conventional pairings of *expression* and *content*. The expressions can be spoken (oral), signed (manual-brachial) or written and the denoted concepts (or “uses” in a more action-oriented Wittgensteinian approach) are *commonly known* to those who are fluent in the language (see Itkonen 1978, this volume; Zlatev 2007, in press).

Given these provisional definitions, we can reformulate the question in the opening sentence as follows: is intersubjectivity a prerequisite for the learning and use of language or is language a prerequisite for understanding others’ (and perhaps even one’s own) “mind”? Thus phrased, the question is a classic dilemma in the literature on “theory of mind”, with arguments in favor for each side of the dependence relation.¹ On the one hand, Bloom (2000: 2) argues persuasively that “... it is impossible to explain how children learn the meanings of a word without understanding of certain non-linguistic mental capacities, including how children think about the minds of others”. There is, indeed, strong evidence that children understand much about adults’ (visual) attention and communicative intentions prior to 18 months (Baldwin 1993, 1993; Tomasello 1999, 2003), and that such understanding appears pivotal for word learning. An often reported result is the following: a child in his second year of life is given a novel toy A to play with, while another toy B is placed out of view. As he is playing with toy A, the experimenter looks at toy B and says: “It’s an X”. The child looks at the experimenter, follows his gaze and discovers toy B for the first time. Importantly, the child assumes that X is the name of toy B, not toy A that he was playing with when he heard X for

the first time. Such findings are problematic for pure associationist models of word learning (e.g. Plunkett 1998).

On the other hand, there is accumulating evidence that language acquisition itself is a determining factor for the development of certain forms of intersubjectivity, especially those which have been linked to the understanding of (false) *beliefs*. For example, deaf children who are not exposed to signed language at an early age understand others' (false) beliefs significantly later than those with signing parents, or hearing children (Peterson and Siegal 1995). Longitudinal co-relational studies indicate that language development predicts performance in tasks of "theory of mind", but not vice versa (de Villiers and Pyers 1997; Astington and Jenkins 1999). Furthermore, training in sentential complement constructions (with or without mental predicates) significantly improves performance in false belief tasks (de Villiers and Pyers 1997; Hale and Tager-Flusberg 2003; Lohmann and Tomasello 2003) and exposure to discourse involving different perspectives independently enhances false belief understanding (Lohmann and Tomasello 2003).

So is it the chicken (language) or the egg (intersubjectivity) that comes first? In order to resolve this dilemma, we need first to clear up our conceptual dusty corners a bit more... Most importantly, it should be emphasized: *Intersubjectivity* \neq "theory of mind"! Elsewhere (Zlatev in press) I address the different perspectives to social cognition these two concepts apply,² but for present purposes it is important to point out the following three characteristics to the approach to intersubjectivity that is here adopted:

- 1) Intersubjectivity is not a unitary capacity: it involves understanding not only beliefs and other proposition-like entities, but other less explicit forms of consciousness: emotions, attentional foci and intentions (cf. Tomasello et al. 2005).
- 2) Intersubjectivity develops in a stage-like manner (in ontogeny and phylogeny) with "lower" stages serving as prerequisites for "higher" ones (e.g. the relationship between *empathy* and *cognitive empathy*, Preston and de Waal 2002).
- 3) Intersubjectivity is bodily-based: understanding others involves identifying with them on a direct bodily level (Merleau-Ponty 1962; Gallagher 2005, 2007; Gallagher and Hutto this volume; Hobson and Meyers this volume), with recent progress in understanding the neural underpinnings of this capacity (Gallese et al. 2004; Arbib 2005; Barresi and Moore this volume).

These three characteristics allow linking intersubjectivity quite naturally to *bodily mimesis* (Donald 1991, 2001; Zlatev 2005, 2007, in press), in a way first suggested by Zlatev, Persson and Gärdenfors (2005a). In this chapter, I will elaborate on this linkage, showing how *five levels* of the *mimesis hierarchy* – defined in the following section – correspond to *different levels of intersubjectivity*. By looking at recent evidence from primatology, and to a lesser degree child development, I will argue that the first two of these levels/stages of intersubjectivity are shared between human beings and (other) great apes, and are therefore quite clearly pre/non-linguistic. The *last* two are specific for us as a species, but they are also clearly linked to language, and thus "post-mimetic". The most theoretically interesting level is the one in between, which appears to be both pre-linguistic and specific for human beings (in its full form). This is our capacity for *triadic mimesis*, as evidenced in e.g. declarative pointing and iconic gesturing (cf. Zlatev, Persson and Gärdenfors 2005b). I will argue that this social-cognitive capacity is the cradle for another cognitive capacity which appears to be uniquely human: *third-order mentality* (e.g. to see that you see that I see), which on its side is central for our ability to share (semantic) knowledge (cf. Itkonen this volume).

Thus, part of the story to be presented is an "egg-based" solution to the dilemma: intersubjectivity grounds language, which then propels the rocket to higher levels. However, if we inquire about the evolutionary origins of triadic mimesis, it appears likely that it is gestural

communication itself that provided the evolutionary niche for its selection, which, in a sense, brings back the role of the “chicken” (proto-)language to center stage.

2. Bodily mimesis and the mimesis hierarchy

In his influential theory of human evolution, Donald (1991) proposed that a form of cognition crucially based on *mimesis*, and a corresponding culture based on mimetic skills such as tool use imitation, ritual dance and gestural communication mediated between the “episodic” cognition of the common ape-human ancestor and the evolution of language as a dominant mode of human communication (see also Hutto this volume). Mimetic representations are according to Donald “conscious, self-initiated, representational acts that are intentional but not linguistic” (ibid: 168). This rather broad definition includes a number of different skills such as imitation, the re-enactment of actions in imagination (and hence planning and rehearsal), and the use of iconic and indexical gestures for intentional communication. Others have suggested a similar “mimetic stage” in ontogeny, but have proposed quite different interpretations of its scope (Nelson 1996; Zlatev 2002, 2003), making it clear that the concept of mimesis requires a more precise definition. Building on Donald’s work, but taking into account some more recent evidence in social neuroscience (see Barresi and Moore this volume) and evidence on the mimetic capacities of non-human primates (Zlatev, Persson and Gärdenfors 2005a), I have in a number of recent publications (Zlatev 2005, 2007, in press) proposed the concept of *bodily mimesis*, which can be defined as follows:

Def: A particular bodily act of cognition or communication is an act of **bodily mimesis** if and only if:

- a) It involves a cross-modal mapping between *exteroception* (i.e. perception of the environment, normally dominated by vision) and *proprioception* (perception of one’s own body, normally through kinesthetic sense);
- b) It is under conscious control and corresponds to – either iconically or indexically – to some action, object or event, while at the same time being *differentiated* from it by the subject;
- c) The subject *intends* the act *to stand for* some action, object or event for an addressee (and for the addressee to recognize this intention);
- d) Without the act being conventional-normative, and
- e) Without the act dividing (semi)compositionally into meaningful sub-acts that systematically relate to each other and other similar acts.

This definition allows us to clarify the relationship between bodily mimesis and a number of related phenomena, on the basis of an evolutionary and developmental model referred to as the *mimesis hierarchy*. The model is unashamedly “progressivist”, and defines each successive stage through the attainment of a new semiotic capacity: (b), (c) and the positive versions of (d) and (e), i.e. *with* conventionality-normativity (d-poss) and compositionality (e-poss). At the same time, it is not a classical stage model in the spirit of Piaget, where each consecutive stage brings with it total reorganization, but a “layered model” (Stern 1985) where earlier capacities continue to co-exist with newer ones, which may subsume but not abolish them. The model (in its current version) distinguishes between 5 stages/levels, determined on the basis of the definition of bodily mimesis given above:

1. Proto-mimesis: only (a), e.g. *neonatal mirroring, contagion, mutual gaze* (cf. “primary intersubjectivity”, Trevarthen 1979)

2. Dyadic mimesis: only (a) and (b), e.g. *action imitation, shared attention, mirror self-recognition*
3. Triadic mimesis: only (a), (b) and (c): e.g. *declarative pointing, iconic gestures, full joint attention* (cf. “pantomime”, Arbib 2005)
4. Post-mimesis₁: (a), (b), (c) and (d-poss), (cf. “protolanguage” Bickerton 2003; “protosign” Arbib 2005)
5. Post-mimesis₂: (a), (b), (c), (d-poss) and (e-poss): e.g. *spoken/signed language* (cf. “symbolic reference”, Deacon 1997)

The first stage, *proto-mimesis* is made possible by a cross-modal mapping between one’s felt bodily actions and the observed actions of others.³ This serves as a basis for bodily mimesis and intersubjectivity – but only a basis, since it is neither under (full) conscious control, nor representational. It can nevertheless account for certain forms of social cognition such as emotional and behavioral contagion and neonatal mirroring. It is in many ways similar of the notion of “primary intersubjectivity” and “interaffectivity” in the developmental literature (Trevarthen 1979; Stern 1985).

For level 2, *dyadic mimesis*, a mimetic act needs to be *volitional* and *representational*, as in Donald’s original characterization of mimesis, given earlier. As condition (b) of the definition states, the notion of *representation* is understood in line with Piaget’s (1945) criterion of differentiation between “signifier” and “signified” from the subject’s point of view (cf. Sonesson in press), adding the requirement that the signifier is a bodily act. Piaget’s example of an infant opening and closing her mouth to model the opening and closing of a matchbox would be an example of an *iconic* correspondence. Children’s acts of pointing for themselves in order to help guide their attention (Bates, Camaioni and Volterra 1975) would qualify as *indexical* mimetic acts.

Level 3 is brought about by adding condition (c), which introduces the necessary *triadic* element in order to make bodily mimesis communicative: the representation or sign is intended to be recognized as such by an addressee, along with the communicative intention itself. This introduces a Gricean element of intentional communication (Grice 1957), involving *intentional* attitudes but not *propositional* ones (cf. Hutto this volume). An example of an iconic sign that fulfills all three conditions is the *miming* of eating by pretending to move a spoon to one’s mouth (made behind a glass door) in order to communicate to a colleague a desire to go for lunch. An indexical mimetic sign would be, for example, a paradigmatic form of *declarative pointing* (Brinck 2003).⁴

Condition (d-poss) distinguishes triadic mimesis from *post-mimesis*, in which the communicative representations are conventional (i.e. commonly known) and normative (i.e. their application is governed by criteria of correctness, of which the users are at least to some degree aware). This qualifies them as being *symbolic*, though not in the sense of Deacon (1997) who insists above all on the property mentioned in (e), the presence of systematic semantic and grammatical relations between the symbols as a definitional criterion for “symbolic reference”. In previous work (Zlatev 2003, 2005) I combined these criteria and did not distinguish between the corresponding two levels of post-mimesis. While it is possible that conventionality and systematicity always come in tandem (in both evolution and ontogeny) I now agree (e.g. with Sonesson in press) that we should not make this an inherent characteristic of the definition of symbolicity. It is at least possible that the “one-word stage” in childhood, and the linguistic skills of language taught apes such as Kanzi (Savage-Rumbough and Lewin

1994) qualify as a “protolanguage”, as suggested by Bickerton (1990, 2003), i.e. as an inventory of symbols, but with very little knowledge of their interrelations. The more “holophrastic” version of this hypothesis – according to which the first expressions are/were more like formulaic sentences such as *How are you?* than words (Zlatev 1997; Wray 2000) – has been referred to as “protosign” (Arbib 2003; 2005). The two possibilities are not exclusive and it seems that some early expressions are more “wordlike” while others are more formulaic, and it is well-known in the child language literature that some children vary in their preference for the two strategies (Nelson 1985).

It is the pervasive grammatical and semantic *systematicity* of all human languages – that has traditionally attracted most of the attention of philosophers and linguists – that brings about the final level: Post-mimesis₂. However, as the model implies, this level would be impossible if it did not, at least in part, rest on the earlier levels.⁵ Nelson and Shaw’s (2002) poignant definition of language as a “socially shared symbolic system” implies this as well: language can indeed be called a “system”, but if it were not for socially shared symbolism, it would not be language.

3. The mimesis hierarchy and levels of intersubjectivity

What gives us ground to apply the mimesis hierarchy to the development and evolution of intersubjectivity? As suggested in the introduction: above all the approach arising from the combination of phenomenology and neuroscience that links social cognition and embodiment, perhaps best summarized by Gallagher (2005, 2007). It was originally Husserl, who over one hundred years ago first criticized the intellectualist perspective on the “understanding of other minds” as a matter of inference or analogy from the knowledge of one’s own mind. As summarized by Gallagher (2007): “For Husserl, understanding another person is not a matter of intellectual inference, but a matter of sensory activations that are unified in or by the animate organism or lived body that is perceiving another animate organism.” This perspective on intersubjectivity was further developed by Merleau-Ponty (1962), in particular through the notion of the “corporeal schema” which serves as “a normal means of knowing other bodies” (Merleau-Ponty 2003: 218). This was originally stated long before neuroscience corroborated the role of the body in understanding others through various mirror neuron systems (Rizzolatti et al. 1996; Gallese, Keysers and Rizzolatti 2004; Arbib 2005), which mediate between the perception of another and the subject’s own proprioception and action. Gallagher’s (2005, 2007) major contribution to this discussion has been to elaborate the distinction between *body schema* and *body image* where the first is pre-conscious and serves as a precondition and backdrop for intentionality, while the latter is “a (sometimes conscious) system of perceptions, attitudes, beliefs and dispositions pertaining to one’s own body” (Gallagher 2005: 37). While one may still question this distinction, and require a finer division within the latter – the *perceptions* of the body are quite different from the *beliefs and attitudes* towards it – I will in general accept it, and reply on it in the discussion of the first two levels of intersubjectivity. Within this “embodied” and action-oriented perspective on the understanding of others as well as the self, bodily mimesis becomes clearly relevant. A central question in relating the mimesis hierarchy and intersubjectivity is whether the respective level serves as a *precondition* and a causal factor for the development of corresponding skills of intersubjectivity. Or is it rather that independently reached insights into the minds of others makes increasingly complex forms of bodily mimesis possible? The problem for the latter scenario is that one would need to account for the emergence of “theory of mind” skills (or “modules”) *independently*, which remains a problem for both evolutionary and child psychology (cf. Hutto in press, this volume). Furthermore, from a Vygotskian perspective stating the priority of the “interpersonal” to the “intra-personal” in cultural development, one would expect that bodily mi-

mimesis to exercise (in ontogeny) and provide selection pressures (in phylogeny) for the development of more refined skills in “mind reading”. At the same time, as with the chicken-and-egg relationship between intersubjectivity and language, we will see that the causality is not unidirectional, and the relationship between mimesis and intersubjectivity may be more pertinently described in terms of *co-evolution*.

3.1 Proto-mimesis

According to the definition of bodily mimesis we can regard some of the most basic forms of intersubjectivity as proto-mimetic to the extent that they consist of interpersonal interactions that involve cross-modal mapping between proprioception and the (visual) perception of others, but lack the characteristics *volition* and *differentiation*. In terms of the distinction between *body schema* and *body image* (Gallagher 2005) mentioned above, proto-mimesis can be said to involve (above all) the body schema, which is largely innate (in the sense of being present at birth) and pre-conscious, rather than the body image, which is gradually constructed with experience and accessible to consciousness. Proto-mimetic forms of intersubjectivity do not require a conceptual differentiation between self and other, which is necessary for establishing a correspondence relation between them. This is not to say that the young infant lives in a completely undifferentiated world in which there is no awareness of self whatsoever, as pointed out by Stern (1985). Nevertheless, even a modern developmental psychologist who emphasizes the role of the awareness of others’ attention and the presence of affective self-consciousness in the first months of life points out that “older infants reveal a greater focus on the self and the younger ones reveal a *more immersed, less detached* focus on the other” (Reddy 2003: 401). This “more immersed, less detached” quality of the earliest forms of intersubjectivity motivates the classification of them as proto-mimetic.

Can this analysis be extended to the (early) interpersonal relations among apes? Neonatal mirroring has also been observed, and appears to be common in chimpanzees (Myowa-Yamakoshi et al 2001, 2004). Since this is typically attributed to a form of *identification* with the person imitated when children are concerned (Meltzoff and Gopnik 1993; Gallagher 2005), it can be viewed as evidence that at least chimpanzees, and possibly other apes and non-human primates possess the capacity for basic proto-mimetic intersubjectivity. The function of such “mirroring” can be related to what is possibly the most basic form of intersubjectivity, both ontogenetically and phylogenetically: the ability to share emotions, or *empathy* (*Einfühlung*). As a proto-mimetic, non-representational capacity, this is testified in early infancy and sometimes referred to as *interaffectivity* (Stern 1985). The well-known experiments described by Trevarthen (1992) show that parent-infant interactions in the first few months take the form of a reciprocal rhythmic “dance”, and that frustration follows if this “attunement” is disrupted (notice the musical metaphors). The suggestion is that emotions such as joy and suffering are perceived *directly*, possibly involving mirror-neuron structures similar to those involved in action recognition and imitation, rather than involving inferences to underlying states (Gallese, Keysers and Rizzolati 2004).

Preston and de Waal (2002) have argued persuasively that as a basic biological mechanism involving the linkage of perception and action, empathy is available to most if not all mammal species. Defining empathy as “any process where the attended perception of the object’s state generates a state in the subject that is more applicable to the object’s state or situation than to the subject’s own prior state or situation” (ibid: 4), they see a clear evolutionary motivation for its emergence in the ability to recognize and understand the behavior of con-specifics. It is an open question how much of such matching between the visually perceived body of the other and the proprioceptively perceived body of oneself is domain-general – and thus can be expected to be general across species – and how much is specialized in the form of species-specific communicative signals such as facial expressions. It is characteristic that signals such

as the famous “play-face” expression of great apes (an evolutionary precursor to the human smile) typically carry emotional rather than referential meaning.

A second mode of intersubjectivity that appears to be of a proto-mimetic nature, at least in human children, is attention. Reddy (2000, 2003, 2005) has argued that prior to awareness of the other’s attention to an external object and much prior to *joint attention* appearing around 12 months (see below), children “show an awareness of others as attending beings, as well as an awareness of self as an object of others’ attention” (Reddy 2003: 357), displayed in phenomena such as eye-contact, intense smiling, coyness, “calling” vocalizations, showing-off etc. Since awareness of self (at this stage) is largely proprioceptive, while awareness of the attention of others (in seeing children) is mostly based on vision, this satisfies the first criterion for bodily mimesis. Reddy’s claim is that such dyadic (though not *dyadic-mimetic*) interactions underpin later developments in intersubjectivity. Evidence for this is the observation that autistic children show difficulties even with such simple interpersonal engagements, “suggesting that whatever is going on in dyadic attentional engagements may indeed be critical, not just as a source of information and experience about attentional behavior, or as a scaffold for the subsequent development of awareness of attention, but also as evidence of awareness of attention” (Reddy 2005: 95).

Until recently it has not been clear whether such awareness of another’s attention exists in the interaction between infant apes and their mothers, but in a recent study, Bard et al. (2005) report that the rates of mutual *gaze* between infants and their mothers are virtually the same in 3-month old human children and 3-month old chimpanzees; 18-20 and 17 times per hour, respectively (though humans tend to engage in much longer bouts of mutual gaze). Furthermore, the authors noticed a “cultural” difference between the apes at Primate Research Institute, Japan and those at Yerkes National Primate Research Center, USA, with the ones in Japan engaging in mutual gaze at much higher rates (22 vs. 12 times per hour), while the ape mothers in the USA *cradled* their infants more often (71% vs. 40% of the total time). Intriguingly, a similar inverse correlation between visual and tactile contact has been observed in human societies, with traditional cultures favoring touch and Western ones gaze: “With reduced physical contact found in Western societies, mutual engagement shifts to the visual system, arguably a more evolutionarily derived pattern” (Bard et al. 2005: 621). Since apes do not seem to differ from humans in the capacity to perform this shift, and possibly even transmit it culturally to their descendents, this confirms the conclusions from neonatal imitation that there is no qualitative difference between the species on the proto-mimetic level.

3.2 *Dyadic mimesis*

In the previous discussion, I proposed that proto-mimetic intersubjectivity, without a clear differentiation between self and other is based on the (mostly) unattended body schema and similarly unattended mechanisms for “body copying”. On the other hand, *dyadic mimetic intersubjectivity* is based on the conscious control of the movements of one’s body and attention to their correspondence to the body of another, whereby one can *imagine* what the other experiences on the basis of one’s own experiences in similar circumstances. In the terminology of Gallagher (2005), I propose that the role that bodily mimesis proper plays for the development of intersubjectivity implies not the body schema but the *body image*. While the body schema and the body image normally interact, Gallagher (2005) shows how in certain pathologies they can be disassociated. Below, I will describe how understanding others’ emotions, attention and intentions can be seen as intimately related to dyadic mimesis.

Whereas (simple) empathy is proto-mimetic, what Preston and de Waal (2002) call *cognitive empathy* requires a differentiation between subject and object where “the subject is thought to use perspective-taking processes to imagine or project into the place of the object” (ibid: 18). Evidence that this is not an isolated phenomenon, but shows a more advanced level of inten-

tionality is the fact that cognitive empathy “appears to emerge developmentally and phylogenetically with other ‘markers of mind’ ... including perspective taking ..., mirror self-recognition ..., deception, and tool-use.” (ibid: 18). Research concerning cognitive empathy in apes has focused on their consolation behavior, which is well-attested in at least chimpanzees, but has not been found in monkey species (de Waal and Aureli 1996) or any other mammalian species. Consolation is cognitively more complex than simple empathy since the consoling individual not only feels that somebody else experiences a particular negative emotion, but also intends to help relieve this, implying an ability to imagine the more positive emotional state.

This supports the interpretation that cognitive empathy involves a more sophisticated representational capacity than what is necessary for simple empathy. Since dyadic mimesis involves both the ability to identify with the other, and at the same time to differentiate between self and other, a natural hypothesis is that it is dyadic mimesis, implicated in e.g. imitation (Zlatev, Persson and Gärdenfors 2005a) that scaffolds the development of such representational capacity (cf. Hutto this volume for a similar proposal).

Since dyadic mimesis allows to “place oneself in the shoes of others”, it also gives the opportunity to understand what someone else is *attending* to. Such “second-order attention” is well testified among great apes (Hare et al. 2000). When two individuals become aware that both are attending to the same object, what results is *shared attention*. This comes a good deal towards the construction of a “consensual reality”, but does not quite reach it. To make a given object X fully intersubjective between you and me, I would need not only to “see that you see X”, (second-order attention, see Figure 1a), but also “to see that you see that I see X” (third-order attention, see Figure 1b) and vice versa – which is one interpretation of what it means to engage in *joint attention*.⁶ Full joint attention thus involves *third-order mentality* and possibly because of this (see below) appears to be beyond the cognitive capacities of apes (Tomasello 1999). It also goes beyond dyadic mimesis, so I will leave this capacity for the time being, but return to it in the following subsection.

INSERT FIGURE 1A AND FIGURE 1B

Concerning the understanding of another’s intentions, it was the received view until the end of the last century that apes cannot do this (e.g. Tomasello 1999). Recently, however, there has been mounting evidence that (at least) “[c]himpanzees understand psychological states – the question is which ones and to what extent”, which is the title of Tomasello, Call and Hare (2003). A wealth of experiments supports this claim. For example, a subordinate and a dominant chimpanzee compete for food placed on the subordinate’s side of two barriers, so that in some cases only the subordinate, but not the dominant can see the food and monitor the visual access of her competitor. The results showed that the subordinates preferentially retrieved the food that dominants could *not* see (and had not seen in the past), implying that chimpanzees are aware of the perceptual states of con-specifics. Together with awareness of the competitor’s goal (i.e. to obtain the food) this allows the “prediction” of the other’s actions and acting accordingly. A dyadic-mimetic interpretation of these facts is that such an understanding can be obtained through the “projection” of one’s own perceptual and motivational state onto the other (“I would get the food if I were in her place!”) at the same time as distinguishing between the self and the other, and does *not* require explicit (propositional) reasoning or inference. There is mounting evidence that even non-enculturated and language-naïve apes are capable of such mimetic enactments.⁷

At the same time, it has not been shown that (non-enculturated) apes are capable of understanding another’s mental states about their own mental states, which would involve, as with joint attention, third-order mentality. In natural settings, some cases of deception *may* be interpreted in a way to involve third-order mentality, but do not require this. For example de

Waal (1998 [1982]) describes the chimpanzee Yeroen who has had a fight with Nikkie, and continues to fake a limp only when in the presence of Nikkie, apparently in an act of wishing to provoke Nikkie's empathy: "I wish to make you see that I hurt". The more parsimonious explanation, however, is that Yeroen has learned from previous experience that he is not bothered by Nikkie when he is hurt: "He may have learned from incidents in the past in which he is seriously wounded that his rival was less hard on him during periods when he was (of necessity) limping" (ibid: 35-36), so he *mimes* the appropriate behavior. Here we have a clear correspondence between dyadic mimesis and second-order intersubjectivity. Notice that Yeroen's limping was not a form of intentional communication – obviously he did not wish Nikkie to understand that he was *faking* a limp – if he did, that would be a case of triadic mimesis.

The conclusion that can be drawn from these various examples is that wild apes as well as those who are exposed to different degrees of human contact (captive, nursery-raised and laboratory-trained apes), but are *not* raised in a "something like a human cultural environment" and thus *enculturated* (Call and Tomasello 1996) can indeed understand second-order mentality. However, such apes do not seem able to understand third-order mentality – neither in the domains of emotion, attention nor intention – in which their own mental state needs to be either intentionally communicated – in collaboration, or hidden – in competition. This corresponds well with the capacity of apes for dyadic mimesis, but their relative difficulty with triadic mimesis, as shown below.

3.3 Triadic mimesis

In the case of triadic mimesis, there is not only an understanding of representational relation between one's bodily motion and object, action or event it corresponds to, but an understanding that such a representational relation can be used *communicatively*. This requires some understanding that the representation (sign) has *the same meaning for the addressee as for the sender*. This involves at least second-order mentality, which was shown above to correlate with dyadic mimesis. But "having the same meaning" is a reflexive notion and implies at least some degree of third-order mentality (see Itkonen this volume). Consider the simple example of what knowing the meaning of the word *cat* implies:

- (1) I know that *cat* means "a small furry animal that meows".
- (2) I expect you know that *cat* means "a small furry animal that meows".
- (3) I expect that you know that I know that *cat* means "a small furry animal that meows".

While it is possible for intentional communication to begin without a full realization of (3), it is practically inevitable that discursive experience (including failures in communication) will promote the development of third-order mentality. Therefore it is possible that *triadic mimesis was one of the major driving forces behind the development of intersubjectivity in hominid evolution*. Unlike competing hypotheses related to "Machiavellian intelligence" (Byrne and Whiten 1988), this puts the focus on *cooperation* rather than competition (see also Brinck and Gärdenfors 2003; Tomasello et al. 2005). A prediction from this hypothesis is that enculturated apes – and these have all been taught at least some degree of sign use – will develop higher-level skills of intersubjectivity. There is support for this prediction. In summarizing some 200 studies of the role of human influence, Call and Tomasello (1996) conclude that "[t]he sociocognitive domains in which humans seem to have the highest effect on apes are intentional communication and social learning" (ibid: 391).

As pointed out earlier, wild apes do not seem to be capable of engaging in full, third order joint attention. Furthermore, as Tomasello (1999: 21) points out, wild apes (a) do not *point to* objects; (b) do not *hold up* objects to show them to others; (c) do not *take someone along* to a place to show them something; (d) do not actively *offer* something to someone; and (e) do not

intentionally *teach* other individuals new behaviors. Tomasello's original account of these absences was based on the claim that apes are unable to understand another's intentions. Given the more recent evidence, this explanation is no longer tenable, and indeed Tomasello et al. (2005) suggest instead that the crucial difference between apes and humans involves the motivation to participate in joint collaborative engagements, and the lack of this motivation prevents apes from constructing "dialogical cognitive representations".

The explanation I propose is similar but more specific: non-enculturated apes fail to develop and understanding of triadic mimesis, and related to that, the ability to engage in third-order mentality. The motivational difference between apes and humans appealed to by Tomasello et al. (2005) cannot be the full explanation since enculturated apes such as Koko, Kanzi and Chantek manage at least (a) and apparently communicative skills (b-e) above as well (Miles 1999), even if in restricted forms. This seems to imply that the human cultural environments⁸ of the enculturated apes have taught them the basics of intentional, sign-mediated communication, and thereby (the roots of) third-order mentality.

How this could occur can be seen again with respect to joint attention, which can be seen as emerging from second-order attention combined with the recognition of another's intention concerning my attention: "I see that you see X" (second-order attention, Figure 1a), and furthermore "I realize that you want me to look at X" (Brinck 2001). In other words, joint attention can be brought forth by understanding a simple form of communicative intention, combined with already existing second-order attention. Thus, *communicating the intention to jointly attend may be said to involve the simplest kind of triadic mimesis*: whatever kind of behavior that is used to convey that intention – some form of index (see the example in the next paragraph) – can be said to *stand for* that intention for both sender and interpreter.

Without enculturation, experiments indicate that apes do not understand communicative intentions. A rather typical example is an experiment by Tomasello, Call and Gluckman (1997), where the authors in different ways indicated for both chimpanzees and two- to three-year-old children which out of three containers contained a reward: by pointing to the correct container; by placing a marker on top of the correct container; and holding up a replica of the correct container. Tomasello (1999: 102) summarizes the results of the experiment as follows:

Children already knew about pointing, but they did not know about using markers and replicas as communicative signs. They nevertheless used these novel signs very effectively to find the reward. In contrast, no ape was able to do this for any of the *communicative signs* that they did not know before the experiment. One explanation of these results is that the *apes were not able to understand that the human beings had intentions toward their own attentional states*. The apes thus treated the communicative attempts of the human as discriminative cues on par with all other types of discriminative cues that have to be laboriously learned over repeated experiences. The children, in contrast, treated each communicative attempt as an expression of the adult's intention to direct their attention in ways relevant to the current situation. [my emphasis]

In other words, while the children clearly understood the communicative intentions of the experimenter, the apes did not. This interpretation is supported by a similar experiment designed to test "false beliefs" (Call and Tomasello 1999), in which the enculturated and language-taught orangutan Chantek clearly performed differently from all the other apes in understanding a human communicator's signals. Even though this was not the goal of the experiment, and Chantek did not score better than the other apes in the false beliefs task, his much better performance could be explained by considering that he understood the signals as communicative signs (in this case indexes), rather than as "discriminative cues".

Finally, we can consider the case of captive apes living in a zoo and thus involved with at least some degree of interaction with human culture. In their study of spontaneous gestural

communication in a group of gorillas in the San Francisco zoo, Tanner and Byrne (1996, 1999), found a wealth of gestures used by several members of the group, in particular by the adult male Kubie, some of which seemed to be used in a communicative way so that:

[w]hether the receiving partner was a human or another ape, the signaling ape made sure that visual contact was established (except for tactile close gestures), and seemed to *understand both the other's potential actions and what the partner might, in turn, understand from his (the signaler's) performance of gestures.* (Tanner and Byrne 1999: 231, [my emphasis])

We can conclude that triadic mimetic intersubjectivity, involving understanding not only of con-specifics' intentions, but their communicative intentions, and consequently a degree of third-order mentality, appears to be *difficult but not completely beyond the cognitive potential of apes*. To realize this potential, apes need an environment that is rich in opportunities for developing triadic mimesis, i.e. a particular form of enculturation. Thus triadic mimesis may be said to be within apes' "Zone of Proximal Development" (ZPD), albeit in its periphery...⁹ If "enculturation" provides the ZPD for present-day apes, it is reasonable to suppose that it did the same for some particularly social group of hominids through a form of "self-domestication" giving rise to a bootstrapping spiral of sign use and intersubjectivity. In the terms of Donald (2001), triadic mimesis must have been within the common ancestor's "zone of proximal evolution", and is therefore a likely candidate for constituting "the missing link" in human cognitive evolution.

3.4 Post-mimesis

What differentiates post-mimetic, or symbolic, cognition from mimesis is the use of *fully conventional* signs, interrelated within a *system* (Deacon 1997; Zlatev 2003). The most obvious example of post-mimesis, involving all the previous features but also symbolicity is a conventional, institutionalized signed language such as ASL (Stokoe 1960) or Swedish Sign Language (Ahlgren 2003). What is the relation between acquiring such a system and intersubjectivity? Prior to addressing this question, let us make the distinction between post-mimesis-1 (protolanguage) and post-mimesis-2 (language), pointed out in Section 2, where only the latter has (extensive) systematicity. A case can be made for apes acquiring the first but hardly the second...

3.4.1. Post-mimesis₁: Protolanguage

Evidence from four of the most successful projects involving the teaching of language to great apes – the chimpanzee (*Pan troglodytes*) Washoe (Fouts 1972, 1973), the gorillas Koko and Michael (Patterson 1978, 1980) and the bonobos (*Pan paniscus*) Kanzi and Panbanisha (Savage-Rumbaugh and Lewin 1994; Savage-Rumbaugh et al. 1998) and the orangutan Chantek (Miles 1990) – has shown that some of the characteristics of language are within the grasp of our nearest non-human relatives. As with children a precondition for the success of these projects has been a cultural environment rich with intersubjectivity and a variety of activities to stimulate communication (Miles 1990). The "ape language" literature contains rather convincing evidence that apes can:

- comprehend the referential (representational) function of spoken words, ASL signs and visual lexigrams, and combinations of these;
- use the sign-tokens in the absence of their referents, i.e. "displacement" (Hockett 1960);
- acquire a considerable vocabulary of words/signs, according to some measurements extending 600 signs, but even according to the most conservative criteria no less than 140 signs
- understand novel combinations of spoken or signed words;

- produce novel combinations of signs;

The following have also been reported, but are considerably less well documented:

- apes can regard the acquired signs as conventional-normative (consensual), to the point of correcting their teachers if the latter do not use these appropriately;
- apes can use language for a number of different functions (speech acts), including labeling, answering, expressing emotion, arguing and insulting;
- apes can use language not only for communication, but for thinking (private speech).

It is therefore possible to agree with Miles (1999: 204), that all great apes “have the intelligence for a rudimentary, referential, generalizable, imitative, displaceable, symbol system” – but with an emphasis on *rudimentary*. It has, for example, not yet been clearly demonstrated whether the spoken or signed utterances of apes conform to consistent principles of grammatical organization. Greenfield and Savage-Rumbaugh (1990, 1991) describe two ordering “rules” in the two symbol combinations of Kanzi, but the data show at best a weak statistical correlation between preferred order and semantic (communicative) function. The most plausible conclusion to the prolonged “ape language” debate therefore seems to be a tie between the extreme proponents and opponents: apes such as Koko and Kanzi can be said to have acquired a form of “proto-language”, which is different from both mimesis due to conventionality, and full language – due to a lack of systematicity which, on its part, is necessary for the production of *narratives*.

But can it even be truthfully said that Koko and Kanzi have acquired semantic conventions? A convention (Lewis 1969; Clark 1996) or a norm (Itkonen 1978) exists as a form of *mutual* (or *common*) *knowledge* among the members of the group that share the convention. A common explication of mutual knowledge is that it consists of *third-order* knowledge: “I know that you know that I know X” (Itkonen 1978, this volume). If this “knowledge” must be in explicit propositional form, then it is unlikely that we can attribute it to the language-taught apes, making it dubious whether we could even call their communicative acquisitions proto-language.

However, is it even warranted to make this requirement when it comes to children? While I earlier argued that triadic mimesis is connected to third-order *mentality*, it is not necessary that the understanding on all three orders is explicit enough to be a matter of *belief*, i.e. a propositional representation that is actively *held* to be true. Consider again the three orders of knowing the conventional meaning of the word *cat* given as (3) above and repeated for convenience:

(3) I expect that you know that I know that *cat* means “a small furry animal that meows”. The highest order thought, *my expectation that you know that I know*, is not a belief for the 4 year old child, since it is taken for granted, without pondering on whether it is true or not. For younger children, even the second order thought is unlikely to be propositional, as evidenced by their inability to understand beliefs proper – false or otherwise. Perhaps it best to call the most basic form of shared cultural knowledge a sharing of *expectations*: we both expect each of us to behave in a certain way given certain conditions (e.g. a red light) and are, however dimly, both aware that this expectation is mutual, and thereby binding. Given this, (3) can be reformulated in a Wittgensteinian manner into (4):

(4) I expect that we are both using *cat* to mean “a small furry animal that meows”

If we are prepared to attribute a degree of semantic knowledge to 2-3 year old children in terms such as (4), then I doubt whether we have good, non-anthropocentric reasons *not* to do

likewise with Chantek and Kanzi. Furthermore, as mentioned earlier, Chantek's experiences with protolanguage obviously bootstrapped his understanding of communicative intentions and third-order mentality, even if he, as all apes so far, was not able to pass a false-belief task (Call and Tomasello 1999).

3.4.2 *Post-mimesis₂: Language*

Irrespective of modality – spoken, signed or written – language is characterized by an unprecedented in animal communication combinatorialness. Recent studies of the spontaneous emergence of Nicaraguan Sign Language (NSL) during the past 25 years show that signed languages have their origin in (triadic) mimesis, but quickly acquire the properties of conventionality and systematicity. Senghas, Kita and Özyürek (2004) compared the co-speech gestures of Nicaraguan speakers of Spanish, with the signing of three “cohorts”, or generations, of learners of NSL and could document some aspects of this transition in detail:

The movements of the hands and body in the sign language are clearly derived from a gestural source. Nonetheless, the analyses reveal a qualitative difference between gesturing and signing. In gesture, manner and path were integrated by expressing them simultaneously and holistically, the way they occur in the motion [event] itself. Despite this analogue, holistic nature of the gesturing that surrounded them, the first cohort of children, who started building NSL in the late 1970s, evidently introduced the possibility of dissecting out manner and path and assembling them into a sequence of elemental units. As second and third cohorts learned the language in the mid 1980s and 1990s, they rapidly made this segmented, sequenced construction the preferred means of expressing motion events. NSL thus quickly acquired *the discrete, combinatorial nature that is hallmark of language*. [my emphasis]

Given their mimetic-gestural origin signed languages have a much greater degree of iconicity than spoken languages and it has been proposed that this plays a role in their faster acquisition by (deaf) children (Brown 1977). Recent studies have questioned this, however, since only a minority of the signs of signed language have transparent iconic meanings, and in a study of 22 children acquiring ASL it was shown that “of the 44 different signs produced by the children before the age of 13 months, 36% were classified as iconic, 30% as metonymic, and 34% as arbitrary” (Bonvillian and Patterson 1999: 253).¹⁰ In this study, the authors compared the rate and pattern of acquisition of ASL by children and that of two gorillas, Koko and Michael. Despite certain differences – the children's acquisition was (unsurprisingly) faster – it was shown that “that the similarities in early development across the species outweigh the differences” (ibid: 260). Thus, it can be concluded that gorillas, and by inference other great apes, not only can acquire the basics of a post-mimetic symbolic system such as ASL, but that they do this in a similar way. Interestingly, however, the gorillas seemed to rely somewhat more on the iconicity of the signs in comparison with the children, so that the proportion of their first 46 signs was somewhat different to the one reported above: 42% iconic, 32% metonymic and 26% arbitrary, while for the first 10 signs this difference was even clearer: 60% iconic, 20% metonymic and 20% arbitrary. This suggests that the apes relied to a greater degree on their skills of triadic mimesis than the children in their acquisition of the sign language, which also would explain why the children quickly progressed beyond the initial level of vocabulary acquisition to learn the systematic character, i.e. the grammar, of the language, while the apes “stagnated” on a simple, protolanguage level.

With the rapid development of grammar and vocabulary around the age of 4, most children also become capable of understanding that others have or lack knowledge or have “false beliefs” (e.g., Perner 1991; Mitchell 1997), implying a *metarepresentational capacity*. It appears that these two developments are closely connected, and that acquiring a language, spoken or

signed, is a major causal factor for developing a fully-fledged “theory of mind”. At least four different sides to language (use) combine to promote metarepresentational capacity.

- Language is a conventional symbolic system, and as such its mastery implies third-order knowledge, which would carry with it training in the understanding of others’ beliefs.
- Two specific (universal) features of human languages are (a) mental predicates such as “think”, “believe”, “know”... and (b) sentential complement constructions such as “say that...”. If one can *meaningfully* formulate sentences such as “I think that you think that X”, then one should be able to think the corresponding *thought*.
- Not just the semantic/grammatical structure of language, but its use in discourse would promote the understanding of others as “mental agents”: There are at least “three kinds of discourse, each of which requires [children] to take the perspective of another person in a way that goes beyond the perspective-taking inherent in comprehending individual linguistic symbols and constructions.” (Tomasello 1999: 173): disagreements, repairs/explanations and meta-discourse.
- Closely related to the above, is the “narrative practice hypothesis” that with linguistic proficiency (usually) comes first apprenticeship and then various degrees of mastery in understanding and producing narratives, through which “children become familiar with both (i) the core structure of folk psychology and (ii) the norm-governed possibilities for wielding it in practice” (Hutto in press).

As pointed out in the introduction, there is accumulating evidence for a strong connection from language to the understanding of beliefs and folk psychology in the case of children (e.g. Peterson and Siegal 1995; de Villiers and Pyers 1997; Astington and Jenkins 1999; Hale and Tager-Flusberg 2003; Lohmann and Tomasello 2003). There is also negative evidence for apes, enculturated or not: as mentioned, Call and Tomasello (1999) used a non-verbal false belief task with chimpanzees and orangutans as well as with human children. While the children’s performance on verbal and nonverbal false belief tasks was highly correlated, supporting the hypothesis of a possible causal connection, none of the apes including Chantek, could pass the nonverbal false belief task even though they succeeded in all of the control trials indicating mastery of the general task demands. A prediction from the present analysis would be that if Chantek, or any of the other “language apes” that have been the subject of so much controversy, could progress in their language development from protolanguage to (systematic and narrative) language, they would also be able to pass false belief tasks. It is indicative that no such evidence has so far been offered.

4. Summary and conclusions

In this chapter, I have argued that there is a close connection between the 5 levels of the evolutionary and developmental model referred to as the *mimesis hierarchy* and corresponding skills in intersubjectivity. There is furthermore a connection between the five levels and what we can call the “type of mentality” involved – reminding that mentality refers to various kinds of states and processes of consciousness, and not only to “propositional attitudes”. These correlations are summarized in Table 1.

Table 1. The mimesis hierarchy, intersubjectivity skills and “types of mentality”

<i>Level</i>	<i>Intersubjectivity skills</i>	<i>Type of mentality</i>
1. Proto-mimesis	(simple) empathy mutual attention	1 st order: lack of complete differentiation between self and other
2. Dyadic mimesis	cognitive empathy shared attention understanding other’s intention (in competitive contexts)	2 nd order: understanding the other through “projection” (identification, but differentiation)
3. Triadic mimesis	joint attention communicative intentions	3 rd order (attention and intentions)
4. Post-mimesis ₁ : protolanguage	semantic conventions	3 rd order (expectations)
5. Post-mimesis ₂ : language	(False) belief understanding	3 rd order and higher (beliefs)

Proto-mimesis is crucially implicated in mutual attention and the awareness of others’ feelings, through a species-general capacity for empathy that has possibly been further developed in the “ultra-social” species *Homo sapiens*. Dyadic mimesis leads to the ability to map between one’s own body and that of others in a more detached, differentiated way, and in this way understand others’ emotions, i.e. cognitive empathy, shared attention and even intentions through a conscious process of “projection”: what would I see/feel/wish if I were you. Unlike earlier claims to the contrary, newer evidence and analyses show that apes do not have much difficulties with this level and that they have the capacity for second-order mentality.

One of the main claims of this chapter is that *the crucial step* in the evolution and development of human intersubjectivity involves triadic mimesis, implying having and understanding others’ communicative intentions. I have argued that this requires third-order mentality: “I want you to do X (e.g. share attention on an object) by recognizing my intention that you do this” from the sender’s perspective and “I understand that you want me to do X” from the recipient’s, but also pointed out that this need not (and does not at this stage) involve beliefs and propositions. Triadic mimesis is obviously difficult for apes to attain, especially in natural conditions. However, through enculturation and especially *through extensive sign use*, some understanding of communicative intentions seems to be within the reach of apes’ “Zone of Proximal Development”, even though it may be in its periphery. Evidence for this is the relative mastery of joint attention by enculturants, and as argued this can be seen to originate in (dyadic mimetic) second-order attention combined with the understanding of the other’s intention that I attend.

Post-mimesis₁ or “protolanguage” implies some understanding of semantic conventions, which I suggested can emerge as shared expectations of common usage, with little if any explicit third-order (propositional) *knowledge*. The long-term studies of language-taught enculturated apes suggest that this level as well, with much persistence, is at least in part accessible to our nearest animal cousins.

Post-mimesis₂, which is identical to language as we know it, has on top of everything else the command of a conventional/normative *system* for communication and *thought*. Arguably it is first with this level the real payoff of using the same system for both meta-functions comes into play, giving us the cognitive benefits of (logical) reasoning, inference, long-term planning etc. that we are so proud of. It gives us, but no other creature on our planet, a metarepre-

sentational capacity, allowing (at least) second-order beliefs, e.g. “I think that you know (or don’t)”.

In summary, bodily mimesis – in its proto, dyadic and triadic forms – can be argued to be a (and possibly *the*) major factor in the evolution of intersubjectivity, with higher mimetic levels bringing along with them more advanced forms of intersubjectivity such as joint attention, and third-order mentality. Sign use itself was suggested to be a driving force in the development of an understanding of third-order mentality, and the performance of enculturants shows that this achievement is within the reach of apes, albeit in special conditions. Therefore one can conclude that (mimetic) sign use was possibly within the “zone of proximal evolution” (Donald 2001) of the common ape-human ancestor, considerably more so than language, characterized by full conventionality and systematicity.

To return to the chicken-and-egg question from the introduction, the intersubjective skills in the first three levels of the mimesis hierarchy (see Table 1) are indeed “pre-linguistic” according to the analysis offered in this chapter, and serve as a *ground* for language – in both evolution and ontogenetic development. However, since they are not “theory of mind modules”, but social skills arising through face-to-face and body-to-body interactions, it would be incorrect to say that *intersubjectivity per se* is a prerequisite for language. It is rather the mimetic “first communions” (Hutto this volume), in which the various skills of intersubjectivity are a natural part – that prepare the way for language. The arrival of the latter marks a major transition, or even two such transitions: to conventionality and systematicity (though it is possible that the two are necessarily linked, Deacon 2003). The understanding of (false) beliefs and folk psychological reasoning are therefore post-mimetic forms of intersubjectivity, since they are based on language, either spoken or signed. This analysis does *not* contradict the claims of those who like Bloom (2000) argue that the acquisition of language presupposes “theory of mind” skills such as joint attention and communicative intentions. What it does contradict is classing such skills as “theory of mind” modules or competencies, since they are triadic mimetic phenomena that are far from being theoretical.

Finally, dividing intersubjectivity along different evolutionary/developmental levels could help resolve even other controversies, such as those voiced by Sinha and Rodríguez (this volume) on the relationship between intersubjectivity and common knowledge. Instead of setting the two in opposition it is quite possible to analyze common knowledge as an advanced, post-mimetic and language-dependent form of intersubjectivity, which I take to be the intention of Itkonen (1978, this volume).

Acknowledgments

This chapter originates from joint work with Tomas Persson and Peter Gärdenfors, which gave rise to the analysis presented by Zlatev, Persson and Gärdenfors (2005a, 2005b), and I thank my former co-authors for the many discussions, including those in which we disagreed. Göran Sonesson, Ingar Brinck, Chris Sinha, Esa Itkonen and Mats Andrén have always been very helpful with feedback related to this and related topics. The Lund University *project Language, Gesture and Pictures in Semiotic Development* and the EU-project *Stages on the Evolution and Development of Sign Use* (SEDSU) provided the funding and the interdisciplinary framework for the type of comparative research that is here described.

References

- Ahlgren, I. 2003. "Teckenspråk." In *Sveriges officiella minoritetsspråk. Finska, meänkieli, samiska, romani, jiddisch och teckenspråk. En kort presentation*. Nordstedt.
- Arbib, M. 2003. "The evolving mirror system: A neural basis for language readiness." In *Language Evolution*, M. Christiansen and S. Kirby 182-200, Oxford: Oxford University Press.
- Arbib, M. 2005. "From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics." *Behavioral and Brain Sciences* 28/2: 105-124.
- Astington, J. W. and Jenkins, J. M. 1999. "A longitudinal study of the relation between language and theory-of-mind development." *Developmental Psychology* 35/5: 1311-1320.
- Baldwin, D. 1991. "Infants' contributions to the achievement of joint reference." *Child Development* 62: 875-890.
- Baldwin, D. 1993. "Infants' ability to consult the speaker for clues to word reference." *Journal of Child Language* 20: 395-418.
- Bard K. A., Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., Costall, A. and Matsuzawa, T. 2005. "Group differences in the mutual gaze of chimpanzees." *Developmental Psychology* 41: 616-624.
- Barresi, J. and Moore, C. this volume. "The neuroscience of social understanding"
- Bates, E., Camaioni, L. and Volterra, V. 1975. "Performatives prior to speech." *Merrill-Palmer Quarterly* 21: 205-226.
- Bickerton, D. 1990. *Language and Species*. Chicago: University of Chicago Press.
- Bloom, P. 2000. *How Children Learn the Meaning of Words*. Cambridge, MA: MIT Press.
- Bonvillian, J. D. and Patterson, F. G. 1999. "Early sign-language acquisition: Comparison between children and gorillas." In *The Mentalities of Gorillas and Orangutans: Comparative Perspectives*, S. T. Parker, R. W. Mitchell and H. L. Miles (eds.), 240-264. Cambridge: Cambridge University Press.
- Brinck, I. 2001. "Attention and the evolution of intentional communication." *Pragmatics and Cognition* 9/2: 255-272.
- Brinck, I. 2003. "The pragmatics of imperative and declarative pointing." *Cognitive Science Quarterly* 3 (4): 429-446.
- Brinck, I. this volume. "The role of intersubjectivity for intentional communication"
- Brinck, I. and Gärdenfors, P. 2003. "Co-operation and communication in apes and humans." *Mind and Language* 18(5): 484-501.
- Brown, R. 1977. "Why are signed languages easier to learn than spoken languages?" In *Proceedings of the National Symposium on Sign Language Research and Training*, W. Stokoe (ed.), 9-24. Silver Spring, MD: National Association of the Deaf.
- Call, J. and Tomasello, M. 1996. "The effect of humans in the cognitive development of apes." In *Reaching into Thought: The Minds of Great Apes*. A. Russon, K. Bard and S. Parker (eds), 371-403. Cambridge: Cambridge University Press.
- Call, J. and Tomasello, M. 1999. "A nonverbal false belief task: The performance of children and great apes." *Child Development* 70/2: 381-395.
- Christiansen, M. and Kirby, S. 2003. (eds.) *Language Evolution*. Oxford: Oxford University Press.
- Clark, H. (1996) *Using Language*. Cambridge: Cambridge University Press.
- de Villiers, J. and Pyers, J. 1997. "Complementing cognition: The relationship between language and theory of mind." In *Proceedings of the 21st Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.
- de Waal, F. (1998 [1982]) *Chimpanzee Politics*. Harper.
- de Waal, F. and Aureli, F. 1996. *Consolation, Reconciliation and a Possible Cognitive Difference Between Macaques and Chimpanzees*. Cambridge: Cambridge University Press.

- Deacon, T. 1997. *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York: Norton.
- Donald, M. 1991. *Origins of the Modern Mind. Three Stages in the Evolution of Culture and Cognition*. Harvard: Harvard University Press.
- Donald, M. 1998. "Mimesis and the Executive Suite: Missing links in language evolution." In *Approaches to the Evolution of Language*, J. Hurford, M. Studdert-Kennedy and C. Knight (eds.), 44-67. Cambridge: Cambridge University Press.
- Donald, M. 2001. *A Mind So Rare. The Evolution of Human Consciousness*. New York: Norton.
- Emery, N.J. 2000. "The eyes have it: the neuroethology, function and evolution of social gaze." *Neuroscience and Biobehavioural Reviews* 24: 581-604.
- Fouts, R. S. 1973. "Acquisition and testing of gestural signs in four young chimpanzees." *Science* 180: 978-980.
- Fouts, R. S. 1972. "Use of guidance in teaching sign language to a chimpanzee (*Pan troglodytes*)." *Journal of Comparative and Physiological Psychology* 80/3: 515-22.
- Gallagher, S. 2005. *How the Body Shapes the Mind*. Oxford: Oxford University Press.
- Gallagher, S. 2007. "Phenomenological and experimental contributions to understanding embodied experience." In *Body, Language and Mind. Vol 1. Embodiment*, T. Ziemke, J. Zlatev and R. Frank (eds.). Berlin: Mouton.
- Gallagher, S. and Hutto, D.D. this volume. "Understanding others through primary interaction and narrative practice"
- Gallese, V., C. Keyners and Rizzolatti, G. 2004. "A unifying view of the basis of social cognition." *Trends in Cognitive Sciences* 8/9: 396 – 403.
- Greenfield, P. M. and Savage-Rumbaugh, E. S. 1991. "Imitation, grammatical development, and the invention of protogrammar." In *Biological and Behavioral Determinants of Language Development*, N. Krasnegor, D. Rumbaugh, M. Studdert-Kennedy and R. Schiefelbusch (eds.), 235-258. Hillsdale, NJ: Erlbaum.
- Greenfield, P. M. and Savage-Rumbaugh, E. S. 1990. "Grammatical combination in *Pan paniscus*: Processes of learning and invention in the evolution and development of language", In *"Language" and Intelligence in Monkeys and Apes*, S T. Parker and K. R. Gibson (eds.), 540-578. Cambridge: Cambridge University Press.
- Grice, P. (1957). "Meaning." *Philosophical Review* 66: 377-88.
- Hale, C. M. and Tager-Flusberg, H. 2003. "The influence of language on theory of mind: A training study." *Developmental Science* 6/3: 346-359.
- Hare, B., Call, J. and Tomasello, M. 2001. "Do chimpanzees know what conspecifics know?" *Animal Behaviour* 61: 139-51.
- Hare, B., Call, J., Agnetta, B. and Tomasello, M. 2000. "Chimpanzees know what conspecifics do and do not see." *Animal Behaviour* 59: 771-85.
- Heyes, C. 1998. "Theory of mind in nonhuman primates." *Behavioral and Brain Sciences* 21: 101-115.
- Hobson, P. and Meyers, J. this volume. "Engaging, sharing and knowing: Foundations for primary intersubjectivity"
- Hockett, C. F. 1960. "The origin of speech." *Scientific American*, 23: 89-96.
- Hutto, D.D. 2004. "The limits of spectatorial folk psychology." *Mind and Language* 19: 548-73.
- Hutto, D.D. this volume. "First communions: Mimetic sharing without theory of mind"
- Hutto, D. in press. *Folk Psychological Narratives: The Socio-cultural Basis of Understanding Reasons*, Cambridge MA: Cambridge University Press.
- Itkonen, E. 1978. *Grammatical Theory and Metascience*. Amsterdam: Benjamins.
- Itkonen, E. this volume. "The central role of normativity in language and linguistics"

- Lewis, D. K. 1969. *Convention: A Philosophical Study*. Cambridge MA: Harvard University Press.
- Lohmann, H. and Tomasello, M. 2003. "The role of language in the development of false belief understanding: A training study." *Child Development* 74/4: 1130-1144.
- Meltzoff, A. and Gopnik, A. 1993. "The role of imitation in understanding other persons and developing a theory of mind." In *Understanding other Minds: Perspectives from Autism*, S. Baron-Cohen, H. Tager-Flusberg and D. J. Cohen (eds.) 335-366. Oxford: Oxford University Press.
- Menyuk, P. and Quill, K. 1985. "Semantic problems in autistic children." In *Communication Problems in Autism* E. Schoper and G. Mesibov (eds.), 127-145. New York: Plenum Press.
- Merleau-Ponty, M. 1962 [1945]. *Phenomenology of Perception*. Routledge and Kegan Paul.
- Merleau-Ponty, M. 2003. *La Nature: Notes, Cours du College de France*. Trans. Robert Valier, Northwestern University Press.
- Miles, H. L. 1990. "The cognitive foundations for reference in a signing orangutan." In "*Language*" and *Intelligence in Monkeys and Apes*, S. T. Parker and K. R. Gibson (eds.). Cambridge: Cambridge University Press.
- Miles, H. L. 1999. "Symbolic communication with and by great apes." In *The Mentalities of Gorillas and Orangutans*, S. Taylor Parker, R.W. Mitchell and H. L. Miles (eds). Cambridge: Cambridge University Press.
- Mitchell, P. 1997. *Introduction to Theory of Mind: Children, Autism and Apes*. London: Arnold.
- Moore, C and Dunham, P. J (eds.) 1995. *Joint Attention: Its Origins and Role in Development*. Hilldale, NJ: Lawrence Erlbaum.
- Myowa-Yamakoshi, M. 2001. "Evolutionary foundation and development of imitation." In *Primate Origins of Human Cognition and Behavior*, T. Matsuzawa (ed.), 349-367. Dordrecht: Springer.
- Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M. and Matsuzawa, T. 2004. "Imitation in neonatal chimpanzees (*Pan troglodytes*)." *Developmental Science* 7/4: 437-42.
- Nelson, K. 1985. *Making Sense. The Acquisition of Shared Meaning*. London: Academic Press.
- Nelson, K. 1996. *Language in Cognitive Development. The Emergence of the Mediated Mind*. Cambridge: Cambridge University Press.
- Nelson, K. and Shaw, L. K. 2002. "Developing a socially shared symbolic system." In *Language, Literacy and Cognitive Development*, J. Byrnes and E. Amseli (eds.), 27-57. Hillsdale, NJ: Lawrence Erlbaum.
- Patterson, F. 1978. "The gestures of a gorilla: Language acquisition in another pongid." *Brain and Language* 5: 72-97.
- Patterson, F. 1980. "Innovative use of language in a gorilla: A case study." In *Children's Language Vol 2*, K. Nelson (ed.), 497-561. New York: Gardner Press.
- Perner, J. 1991. *Understanding the Representational Mind*. Cambridge, MA: MIT Press.
- Perner, J., Leekam, S. and Wimmer, H. 1987. "Three-year-old's difficulty with false belief: the case for a conceptual deficit." *British Journal of Developmental Psychology* 5: 125-137.
- Peterson, C. C. and Siegal, M. 1995. "Deafness, conversation and the theory of mind." *Journal of Child Psychology and Psychiatry and Allied Disciplines* 36: 459-474.
- Piaget, J. (1945) *La Formation du Symbole Chez l'enfant*. Delachaux et Niestlé.
- Plunkett, K. 1998. "Language and connectionism." *Language and Cognitive Processes* 13: 105-127.
- Preston, S. and de Waal, F. 2002. "Empathy: Its ultimate and proximal bases." *Behavioral and Brain Sciences*, 25: 1-72.
- Reddy, V. (2000). "Coyness in early infancy." *Developmental Science* 3 (2): 186-192.

- Reddy, V. 2003. "On being an object of attention: Implications of self-other-consciousness." *Trends in Cognitive Science* 7/9: 397-402.
- Reddy, V. 2005. "Before the "third element": Understanding attention to self." In *Joint Attention*, N. Klein (ed.), 85-109. Oxford: Oxford University Press.
- Rizzolatti, G. and Arbib, M. 1998. "Language within our grasp." *Trends in Neurosciences* 21: 188-194.
- Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. 1996. "Premotor cortex and the recognition of motor actions." *Cognitive Brain Research* 3 : 131-141.
- Saussure, F. de 1916. *Cours de Linguistique Générale* [Course in General Linguistics]. Paris : Payot.
- Savage-Rumbaugh, S., Shanker, S., and Taylor, T. 1998. *Apes, Language and the Human Mind*. Oxford: Oxford University Press.
- Savage-Rumbough, S. and Lewin. R. 1994. *Kanzi: The Ape at the Brink of the Human Mind*. New York: John Wiley.
- Searle, J. 1992. *The Rediscovery of the Mind*. MIT Press.
- Senghas, A., Kita, S. and Özyürek, A. 2004. "Children creating core properties of language: Evidence from an emerging sign language in Nicaragua." *Science* 305: 1779-1782.
- Sinha, C. and Rodríguez. This volume. "Language and the signifying object: From convention to imagination"
- Sonesson, G. 2001. "From semiosis to ecology. On the theory of iconicity and its consequences for the ontology of the Lifeworld." *VISIO* 6/2-3: 85-110.
- Sonesson, G. *in press*. "The *Symbolic Species* revisited. Considerations on the semiotic turn in cognitive science and biology."
- Stern, D. 1985. *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*. New York: Basic Books.
- Stokoe, W. 1960. "Sign language structure: An outline of the visual communication system of the American deaf." *Studies in Linguistics, Occasional Paper 8*, University of Buffalo, New York.
- Tanner, J. E. and Byrne, R. W. 1999. "The development of spontaneous gestural communication in a group of zoo-living lowland gorillas." In *The Mentalities of Gorillas and Orangutans – Comparative perspectives*, S. T. Parker, R. W. Mitchell and H. Lyn Miles (eds.). Cambridge: Cambridge University Press.
- Tanner, J. E. and Byrne, R. W. 1996. "Representation of action through iconic gesture in a captive lowland gorilla." *Current Anthropology* 37/1: 162-73.
- Terrace, H. S., Petitto, L. A., Sanders, R. J. and Bever, T. G. 1981. *Science* 211/4477: 87-88.
- Tomasello, M. 1999. *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. 2003. *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M., Call, J. and Gluckman, A. 1997. "Comprehension of novel communicative signs by apes and human children." *Child Development* 68: 1067-1080.
- Tomasello, M., Call, J. and Hare, B. 2003. "Chimpanzees understand psychological states – the question is which ones and to what extent." *Trends in Cognitive Sciences* 7/4: 153-156.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. 2005. "Understanding and sharing intentions: The origins of cultural cognition." *Behavioral and Brain Sciences*: 28: 675-735.
- Tomasello, M. and Call, J. *in press*. "Do chimpanzees know what others see – or only what they are looking at?" In *Rational Animals*, S. Hurley and M. Nudds (eds.). Oxford: Oxford University Press.

Trevarthen, C. 1979. "Communication and cooperation in early infancy: A description of primary intersubjectivity." In *Before Speech*, M. Bullowa (ed.), 321-347. Cambridge: Cambridge University Press

Trevarthen, C. 1992. "An infant's motives for speaking and thinking in the culture." In *The Dialogical Alternative*, H. Wold (ed.), 99-137. Oslo: Scandinavian University Press.

Vygotsky, L. S. 1962. *Thought and Language*. Cambridge MA: MIT Press.

Vygotsky, L. S. 1978. *Mind in Society. The Development of Higher Psychological Processes*. Harvard University Press.

Wittgenstein, L. 1953. *Philosophical Investigations*. Basil Blackwell.

Woll, B. and Kyle, J. 2004. "Sign language." *Encyclopedia of Language and Linguistics*. Oxford: Elsevier.

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. Hurford (eds.), 285-302. Cambridge: Cambridge University Press.

Zlatev, J. 1997. *Situated Embodiment. Studies in the Emergence of Spatial Meaning*. Stockholm: Gotab.

Zlatev, J. 2002. "Mimesis: The "missing link" between signals and symbols in phylogeny and ontogeny." In *Mimesis, Sign and the Evolution of Language*, A. Pajunen (ed.), 93-122. University of Turku Press.

Zlatev, J. 2003. "Meaning = Life (+ Culture). An outline of a unified biocultural theory of meaning." *Evolution of Communication* 4/2: 253-296.

Zlatev, J. 2005. "What's in a schema? Bodily mimesis and the grounding of language." In *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, B. Hampe (ed.), 323-342. Berlin: Mouton.

Zlatev, J. 2007. "Language, embodiment and mimesis." In *Body, Language and Mind. Vol 1. Embodiment*, T. Ziemke, J. Zlatev and R. Frank (eds.). Berlin: Mouton.

Zlatev, J. in press. "Love Reign O'er Me: The intersubjective grounding of language in bodily mimesis." In *New Directions in Cognitive Linguistics*, V. Evans and S. Pourcel (eds.). Amsterdam: Benjamins.

Zlatev, J., Persson, T. and Gärdenfors, P. 2005a. "Bodily mimesis as "the missing link" in human cognitive evolution." *Lund University Cognitive Studies* 121.

Zlatev, J., Persson, T. and Gärdenfors, P. 2005b. "Triadic bodily mimesis is the difference." Commentary to Tomasello et al. (2005) *Behavioral and Brain Sciences*: 28: 675-735.

Zlatev, J. and Heaton, P (forthcoming) "Mimesis, language and autism"

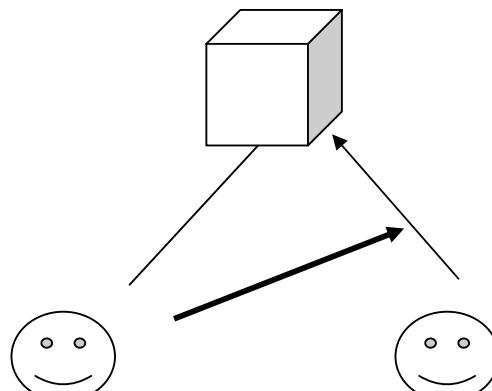


Figure 1a. *Shared attention*: Second-order attention. "I see that you see X" (and vice versa)

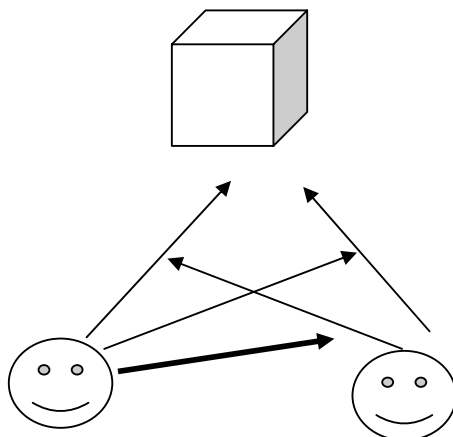


Figure 1b. *Joint attention*: Third-order attention: “I see that you see that I see X”
(and vice versa)

- ¹ For an interesting discussion of these issues, though from a predominantly “theory of mind” perspective, cf. www.interdisciplines.org/coevolution.
- ² See also Sinha and Rodríguez (this volume) and Hutto (this volume) for critiques of overly “mentalist” interpretations of intersubjectivity.
- ³ See Berresi and Moore (this volume) for a detailed neuroscientific model of how first-person and third-person information appears to be “matched” in the posterior parietal cortex.
- ⁴ The iconic, but not the indexical aspects of triadic mimesis are similar to the notion of *pantomime*, constituting Stage 5a in the somewhat similar evolutionary model proposed by Arbib (2003, 2005).
- ⁵ One case in which language acquisition relies (much) less on bodily mimesis is *autism*. Predictably, however, even the language of high-functioning persons with autism displays semantic and pragmatic abnormalities (Menyuk and Quill 1985; Zlatev and Heaton forthcoming).
- ⁶ The terms “shared attention” and “joint attention” have, unfortunately, not been standardized in the literature and are often used interchangeably. One exception is Emery (2000), who however uses the terms in nearly the reverse sense as that adopted here, with “shared attention” being the more higher-order phenomenon.
- ⁷ While it still not conclusively shown that apes are capable of such mental “projection”, and it is conceivable how the evidence can be explained in a more behaviouristic manner involving learning generalizations over other individuals’ behaviour in relation to food, the mental explanation is (a) ultimately more parsimonious (cf. Tomasello and Call in press) and (b) consistent with the performance of human children in roughly comparable stages of development.
- ⁸ Consisting of what Wittgenstein (1953) called the “forms of life”, which provide the necessary context for the emergence and functioning of intentional communication and language.
- ⁹ ZPD is the notion introduced by Vygotsky (1962, 1978) to refer to skills that children could acquire with the help of adults, but not alone.
- ¹⁰ “Metonymic” signs are such that involve some degree of iconicity between the sign and the referent, but “the tie between the sign and its meaning is not readily apparent – one would be unlikely to guess the meaning of a metonymic sign simply seeing it produced” (Bonvillian and Patterson 1999: 252). In the distinction introduced by Sonesson (2001) they involve *secondary* (rather than primary) iconicity, which can be perceived first after knowing the specific sign relationship.