From proto-mimesis to language: Evidence from primatology and social neuroscience

Jordan Zlatev

Lund University, Centre for Languages and Literature, Box 201, 221 00 Lund, Sweden

A R T I C L E   I N F O

Keywords:
Bodily mimesis
Primates
Imitation
Gesture
Mirror neurons
Evolution

A B S T R A C T

How can we reconcile the conception of language as a conventional-normative semiotic system with a perception/action-based account of its structure and meaning? And why should linguistic meaning – as opposed to linguistic expression – be so closely related to motor activity and its neural underpinnings, as suggested by recent findings? A conceptual framework and evolutionary scenario building on the concept of bodily mimesis [Zlatev, J., 2005. What's in a schema? Bodily mimesis and the grounding of language. In: Hampe, B. (Ed.), From Perception to Meaning: Image Schemas in Cognitive Linguistics. Mouton de Gruyter, Berlin, pp. 313–343] imply answers to these questions. The article presents evidence for a particular evolutionary stage model by reviewing recent evidence on the capacity of non-human primates for intersubjectivity, imitation and gestural communication, and from neuroscientific studies of these capacities in monkeys and human subjects. It is argued that “mirror neuron” systems can subserve basic motoric and social capacities, but they need to be considerably extended in order to provide an efficient basis for bodily mimesis, and even more so for language. It is argued that while language may be ultimately “grounded” in perception and action, it is essential not to try to reduce it to them.

© 2008 Published by Elsevier Ltd.
mimesis. Finally, I will return to the two questions posed in the preceding section.

2. Bodily mimesis and the mimesis hierarchy

In his influential theory of the origins of human cognition and culture Donald (1991) argued that a form of cognition crucially based on mimesis, and a corresponding “mimetic culture” characterizing the primitive societies of Homo ergaster/erectus came between the “episodic” cognition of the common ape-human ancestor and the emergence of language. Mimetic representations are defined as “conscious, self-initiated, representational acts that are intentional but not linguistic” (Donald, 1991, p. 168). In this analysis, mimesis involves a number of different skills such as imitation, gestural communication and the re-enactment of actions in imagination, sharing features such as reference, intentionality and autocoercing, made possible by what Donald calls a “mimetic controller”.

While the proposal is generally plausible, there is much in Donald’s concept of mimesis that has remained unclear: what pre-adaptations did mimesis build on, and what exactly differentiates it from, e.g. signed language? How to account for the semiotic differences between, e.g. imitation and gesture, and did they evolve in any particular order? Are all aspects of mimesis under conscious control, or are some “sub-personal”? Which senses of the philosophically loaded (and ambiguous) notions of intentionality and reference are intended in the definition? More empirically, it has been convincingly shown that (even non-enculturated) apes are in fact capable of some forms of imitation, e.g. (Custance et al., 1995; Whiten, 2000; Bjorklund and Bering, 2003), pointing (Leavens and Hopkins, 1999) and possibly even iconic gesture (‘Tanner, 2004). In terms of neuroscience, it is not evident how the “mimetic controller” relates to recent influential notions such as “mirror neurons” (Gallese et al., 1996; Arbib, 2005), “brain simulation” (Berthoz, 2000) and “shared representations” (Decety and Sommerville, 2003). Finally, given the somewhat vague nature of the concept, it has been applied quite differently to the analysis of human ontogeny (Nelson, 1996; Zlatev, 2003), making it difficult to evaluate suggestions that it can help account for a degree of parallelism between phylogeny and ontogeny.

In attempting to resolve these perceived difficulties I have proposed a series of related definitions of the concept of bodily mimesis (Zlatev et al., 2005; Zlatev, 2005, 2007a) which in Zlatev (2008) is formulated as follows:

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 – 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.

(e) Without the act dividing (semi)compositionally into meaningful sub-acts that systematically relate to each other and other similar acts.

This definition clarifies a number of issues on the relationship between bodily mimesis and similar but distinct phenomena. At the same time, it allows the formulation of a model of human evolutionary (and human development) called the mimesis hierarchy. The model defines each successive stage through the clear attainment of a previously unavailable cognitive/semiotic capacity: condition (a) in the definition serves as the foundation, and the capacities listed under (b) and (c) and the positive versions of (d) and (e), i.e. with conventionality/normativity (+d) and with semiotic systematicity (+e), constitute progressively “higher” stages in human socio-cognitive evolution. 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, 2000 [1985]) where earlier capacities continue to co-exist with newer ones, which may subsume but not abolish the earlier ones (Zlatev, 2003, 2008).

Condition (a) states that a special form of active perception (Berthoz, 2000), in which (dynamic) aspects of the environment – especially the actions of conspecifics – are mapped onto one’s own bodily actions and sensations, is a necessary condition for bodily mimesis. In the case of social interactions this gives rise to “shared representations” between self and other (Decety and Sommerville, 2003), or “self-other matching” (Barresi and Moore, 2008). However, this is not sufficient for true mimesis. If only condition (a) is fulfilled, as appears to be the case in contagion and neonatal imitation, the individual is capable of proto-mimesis but not bodily mimesis proper. Condition (b) states that a mimetic act needs to be both volitional and representational, as in Donald’s original definition, and explicates the notion of representation in line with Piaget’s double criterion of both differentiation and correspondence between “signifier” and “signified” from the subject’s point of view (Piaget, 1945; Sonesson, 2007), adding the requirement that the signifier is a bodily act. Piaget’s example of an infant opening its mouth and closing her mouth to model the opening and closing of a matchbox would be an example of an iconic correspondence (i.e. based on similarity) between the act and the object of attention. Children’s acts of pointing for themselves in order to help guide their attention (Bates et al., 1975) would qualify as indexical mimetic acts (i.e. based on spatiotemporal contiguity). Volition and representation are implied in the development of a “sense of a core self” (Stern, 2000 [1985]) in which the body is felt to be “one’s own”, and can be both differentiated from the environment and consciously directed at it, thus fulfilling the definition of a sign or a representation: X stands for Y for S(subject), e.g. mirror image stands for own body for observer, in the case of mirror self-recognition.

However, if condition (c) is not fulfilled, the subject may be involved only in dyadic mimesis. Condition (c) introduces the necessary triadic element in order to make bodily mimesis communicative: the representation or sign is intended to be recognized as such (i.e. understood) by an addressee, as well as the communicative intention itself (Grice, 1989). Thus there needs to be a double differentiation: between self and other and between the “expression pole” of the mimetic act and its “semantic pole” (Langacker, 1987). The three-part relationship between self-initiated mimetic gesture, its meaning and the receiver of the intended meaning is what justifies calling this “triadic” mimesis. An example of an iconic sign that fulfills all three conditions (a)–(c) is the miming of eating by pretending to move a spoon to one’s mouth (e.g. made behind a glass door) in order to communicate to a colleague that it is time for lunch. A mimetic sign that combines iconicity (since the motion and direction of the hand resembles the intended direction

\footnote{Neonatal imitation (or mirroring) does not require a clear differentiation between self and other, while “true” imitation does, as well as differentiating between the means and the goal of an action (see Section 3.2 for further explication of this).}

\footnote{As suggested in Sonesson (2007) I regard the terms representation and sign as synonymous (once properly defined). Since these are arguably the pivotal terms of cognitive science and semiotics, respectively, identifying the two should contribute to the convergence of these two different traditions. Note also that the kind of representations/signs I am arguing for are both bodily-based and dynamic, and thus not prone to the critiques that traditional representationalism has deservedly received in both cognitive science (Varela et al., 1991) and semiotics (Thibault, 2004).}
of attention of the addressee) and indexicality is declarative pointing.

If an act fulfills not only (a)–(c), but also the positive version of condition (d), then it is not properly speaking mimetic, but rather symbolic, since by adding the property of conventionality/normativity, the iconic and/or indexical motivation – or “ground” in Peircean terms (Sonnson, 2006) – of the sign loses much of its function, allowing the relationship between expression and content to become increasingly arbitrary. On this account, arbitrariness and conventionality are related, but non-synonymous notions, thus accommodating the fact that up to 50% of the signs of signed languages can be said to be more or less iconic (Woll and Kyle, 2004), but nonetheless conventional (and thus largely differing among signed languages). A system of communication that has properties (a), (b), (c) and (+d) can be called a “protonlanguage”.

Finally, by introducing the positive version of (e), semiotic systematicity, involving hierarchical relations between composite and “simple” signs (corresponding to what is usually referred to as “compositionality”, without implying any specific mechanism for its implementation), and furthermore relations to other signs (Deacon, 1997), the border between protolanguage and language has been traversed conceptually. In the following sections I will discuss how this could have been done actually in evolution.3

This outline of the model is summarized in Table 1, relating the five different stages to particular communicative and cognitive skills which testify to the attainment of particular novel capacities, and some non-synonymous but related concepts proposed in the literature on social learning and development, some of which are related to similar, though non-equivalent notions used in the literature.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Novel capacity</th>
<th>Cognitive/communicative skills</th>
<th>Related notions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>– Neonatal imitation</td>
<td>– Primary intersubjectivity ([Trevathen, 1979])</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Mutual gaze</td>
<td>– Sense of a core self ([Stern, 2000 [1985]])</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Imitation</td>
<td>– Secondary intersubjectivity ([Trevathen and Hubley, 1978])</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– (cognitive) Empathy</td>
<td>– (true) Imitation ([Tomasello, 1999])</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Shared attention</td>
<td>– Mirror self-recognition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– (full) Joint attention</td>
<td>– Declarative pointing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– One-word utterances</td>
<td>– Holophrases</td>
</tr>
<tr>
<td>Dyadic mimesis</td>
<td>Volition and representation</td>
<td>– Symbolic</td>
<td>– Pantomime ([Arbib, 2005])</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Conventionality/normativity</td>
<td>– Gesture ([Corballis, 2002])</td>
</tr>
<tr>
<td>Triadic mimesis</td>
<td>Communicative intentions</td>
<td>– Protolanguage ([Bickerton, 2003])</td>
<td>– Symbolic reference ([Deacon, 1997])</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Declarative pointing</td>
<td>– Tertiary intersubjectivity ([Halliday, 1993])</td>
</tr>
<tr>
<td>Post-mimesis 1</td>
<td>Conventionality/normativity</td>
<td>– Protosign ([Arbib, 2005])</td>
<td>– Narrative self ([Stern, 2000 [1985]])</td>
</tr>
<tr>
<td>Post-mimesis 2</td>
<td>Semiotic systematicity</td>
<td>– Symbolic</td>
<td></td>
</tr>
</tbody>
</table>

Each stage corresponds to a condition in the definition of bodily mimesis (see text), with the last two stages violating the respective conditions. The different stages are related to similar, though non-equivalent notions used in the literature.

Evidence for the model from studies in primate social cognition and social neuroscience, and point out its evolutionary implications.

But first I need to make one more clarification. Both bodily mimesis and language are manifested in specific acts, which are originally interpersonal and overt, but can be internalized and used for thought (Vygotsky, 1962 [1934]; Place, 1945; Donald, 1991). Such internalized, or covert mimesis can be seen to correspond to what has been described as “action images” (Jeannerod, 1997) or “embodied simulation” (Gallese, 2005). However, as commonly used in the literature (Gallese and Lakoff, 2005) the notion of “simulation” frequently conflates two different sorts of processes: those which involve consciously accessible imagery (Kosslyn et al., 2001) and those which occur on a “sub-personal” level (Barsalou, 1999). I disregard here the question whether the latter phenomenon is properly called simulation (Gallagher, 2005), but insist on the need to distinguish the two senses: consciousness is undeniable based on the activity of the brain, but it is not identical to it (Searle, 1992). To give a simple example: Unconscious processes of “simulation” are going on in my brain as I observe a soccer game, but only if I have a more or less vivid image (either kinesthetic or visual, or a combination of these) of, say, Zidane head-butting Materazzi in the final of the 2006 Football World Cup, can I say to have a mental representation of the event (i.e. differentiated from the event from my point of view), Therefore simulation of the fist kind is (possibly) proto-mimetic, while my imagination of the event is dyadic mimetic, and if I were to re-enact Zidane’s action (on an imaginary opponent) for you, it would be triadic mimetic.5 These distinctions will become even clearer in Section 4 when I review evidence for different neural mechanisms of unconscious “simulation” (or more appropriately, resonance) and those supporting imagination and gesture.

3. The mimesis hierarchy and the evolution of social cognition

In this section, I apply the mimesis hierarchy to evidence from primatology with respect to three domains of social cognition that I have previously argued constitute forms of bodily mimesis: imitation, intersubjectivity, and gestures (Zlatev et al., 2005; Zlatev, 2008). In other words, I will consider simple forms of these capac-

3 It is possible that (+d) and (+e) always come in tandem since there are no existing languages which have only one but not the other, and Deacon has argued that this may be necessary so due to “semiotic constraints” ([Deacon, 2003]). In some previous formulations of the model I have accepted this ([Zlatev, 2003]). However, there are conceptual reasons for separating the two ([Sonnson, 2007] and as I show in this article, there are empirical reasons as well.

4 Notice that this implies taking a perspective: e.g. either that of Zidane or Materazzi – I can imagine myself either head-butting, or falling down and screaming, but not both at the same time. The question of identification is not addressed by appeals to “subpersonal” simulatory processes (Gallese, 2005; Gallese and Lakoff, 2005).
ities that are in essence proto-mimetic, capacities that are dyadic mimetic, those that are triadic mimetic – and whether apes display behaviors that can be even said to involve post-mimesis, in either its more simple or complex forms. Since it turns out that the higher we go in the hierarchy, the more limited the non-human primate capabilities are, and furthermore that apes “score higher” than monkeys, this can be taken as evidence that the model can map, however tentatively, onto an evolutionary sequence. The levels of the hierarchy thus correspond to hypothetical stages in hominid cognitive evolution.⁵

In the study of human cognitive evolution, there seems to be an increasing use of the stage concept, with nearly all more elaborated recent theories involving some sort of a stage model (Donald, 1991; Deacon, 1997; Johansson, 2005; Arbib, 2005). Skeptics sometimes point out that a stage model in evolution seems to imply qualitative transitions, while both the nature of evolutionary processes involving many small mutations/gene recombinations and the fossil evidence seem to speak in favor of a prolonged gradual process. However, there is no contradiction between a stage-based theory and the evolutionary evidence. Firstly, it is possible to have relatively discrete transitions, or “punctuated equilibria” (Eldridge and Gould, 1972) within a basically gradual framework; i.e. without this implying any form of “saltations” or sudden “macro-mutations”. The likelihood for fairly rapid and abrupt changes is even higher when biological and cultural evolution interacts, as has most likely been the case with human cognitive evolution (Deacon, 1997). Furthermore, the combination of paleoontological, neuro-scientific and comparative psychological evidence seems to support qualitative differences at least between monkeys and apes, between apes (and the “common ancestor”) and Homo ergaster/erectus, and between the latter and Homo sapiens. If we furthermore consider the cultural evolution that has proceeded since the appearance of Homo sapiens, we must also take into consideration the difference between cultures without and with organism independent representations (Sonesson, 2007), between “mythic” and “theoretic” cultures (Donald, 1991), or between “protolanguage” and language (Arbib, 2005).

3.1. Proto-mimesis: contagion, neonatal imitation and mutual gaze

Proto-mimesis is displayed in phenomena such as emotional and attential contagion, and neonatal imitation (for characterizations, see below) which involve some form of self-other “mapping”, but are not under volitional control, and do not require a (conscious) distinction between self and other. It can also be argued to characterize the experience of the infant during episodes of mutual gaze.

Various forms of contagion are incontestably established in both monkeys and apes. For example, monkeys (and other social mammals) display emotional distress when a con-specific is in pain (Preston and de Waal, 2002). “Laughter” is as we know from everyday experience highly contagious, and it is also so in apes, as argued by Provine (2000) – who suggests that it has originated through phylogenetic ritualization from the breathing patterns during tumble-play. Yawning contagion was for a long time attested only in humans, but Anderson et al. (2004) and Amici and Call (2007) have shown that at least some chimpanzees yawn as well when presented videos of conspecifics’ yawning (though not of human yawning). On the basis of findings that chimpanzee infants do not yawn contagiously, and that human children have not been attested to do so before 5 years of age, Anderson et al. (2004) suggest that it is based on more sophisticated mechanisms, involving identification and possibly even “theory of mind”. Still, given that (contagious) yawning is a clear example of an “action” performed non-volitionally, I classify it as being proto-mimetic, and in contrast to Anderson et al. (2004) predict that yawning contagion will be found also in monkeys.

The co-orienting response of primates, or even domestic goats (Kaminski et al., 2005), in which an animal turns its attention in the same direction as a con-specific when the latter re-directs its body posture can be regarded as a form of attentional contagion (Brinck, 2008). Whether such behaviors all involve a form of exteroception-propiroception matching is not clear, but they can all be accounted for through a more general kind of (“sub-personal”) perception-action mechanism that makes good evolutionary sense for social animals (Preston and de Waal, 2002). When the process results in a similar emotional state in the original “actor” and in the “resonator” this can be plausibly regarded as a (simple) form of empathy. At the same time, both attentional and emotional contagion can be distinguished from even more basic processes of social coordination such as the schooling of fish, in that while governed by unconscious processes, contagion results in similar perceptual or emotional states of phenomenal consciousness – and thus constitute a basic form of intersubjectivity (Zlatev, 2008) – while fish schooling results only in the coordination of behavior, without (necessarily) involving any shared “intentional object”, to use the term from phenomenology, rather than the more familiar, but also more problematic notion of “mental representation”.

Far from being a distinguishing feature of human infants, neonatal imitation of, e.g. tongue protrusion and mouth opening has been shown to be present in chimpanzees (Myowa-Yamakoshi et al., 2004) and even in monkeys (Ferrari et al., 2006). In the case of chimpanzees this disappears after 2–3 months, similarly as in children, hence suggesting similar neural bases (see Section 4). In macaques its time-window is much shorter, apparently only a few days, and restricted to only a few simple “gestures”: tongue protrusion and mouth opening, showing a possibly homologous, but less developed form of this basic capacity for social interaction in comparison with apes, and even more so human beings.

More controversial is the status of mutual gaze, as a basic form of experience-sharing between infants and mothers (and other care-givers). Human infants participate in extensive bouts of eye-contact with their mothers and this, together with co-occurring vocalizations, have been regarded as a conspicuous and developmentally significant form of “primary intersubjectivity” (Trevathan, 1979; Reddy, 2003) in which a calibration of the (typically positive) affective state of the two takes place. It has been reported (Bard et al., 2005) that the rates of eye-contact episodes between infants and mothers are virtually the same in 3-month-old human children and 3-month-old chimpanzees; 18–20 and 17 times per hour, respectively. Still, human mother–infant dyads were found to engage in much longer bouts of mutual gaze.

In sum, while it is premature to dismiss the possibility for differences between human beings and non-human primates in primary interactions related to proto-mimesis, it seems that these are less pronounced than has been suggested by some (Hobson, 2004). With the exception of the turn-taking in vocalizations that is observed in “proto-conversations”, the other behaviors that seem to testify to the presence of primary intersubjectivity between infants and con-specifics (above all mothers) – contagion, neonatal imitation, and mutual gaze – seem to be present in non-human primates, and above all in our closest non-human relative, the chimpanzees.

---

⁵ This logic does not rest on the faulty premise that monkeys, apes and humans constitute the representatives of actual evolutionary stages, but is based on the assumption that present day monkeys have evolved less from the common monkey-ape ancestor than apes, and apes less from the common ape-human ancestor than humans, which is a plausible assumption given the different species’ relative increase in brain size compared to the common ancestor, (cf. Donald, 1991; Deacon, 1997; Johansson, 2005).
3.2. Dyadic mimesis: imitation and imaginary re-enactment

Unlike proto-mimesis, where “imitation” is highly limited in scope and can plausibly be classified as a form of mimicry (Donald, 1991, 2005), true imitation involves (a) a self/other differentiation and (b) volitional control of one’s body movements, which are both necessary in order to match one’s movements to those of the model in performing a novel action. Furthermore, it has been argued that to engage in “true” imitation, as opposed to emulation (in which an action of with same goal is enacted, but using different means) response facilitation or stimulus enhancement – for definitions and review see (Whiten et al., 2004) – it is necessary to be able to understand the goal of the modeled action, and to see the action as a means to achieve this goal (Tomasello, 1999). While it is fairly well established that monkeys do not imitate (Visalberghi and Fargaszy, 2002), even when one disregards the requirement of goal-understanding, it has recently become clear that great apes are capable of imitation to some degree. This has been shown in tasks of “do-as-I-do” experiments with chimpanzees (Custance et al., 1995) and gorillas (Byrne and Tanner, 2006) “artificial fruits” (Whiten et al., 2004) and deferred imitation with respect to actions on objects (Bjorklund and Bering, 2003). Successful imitation has been recorded both on the “program level” and the “action level”, the first requiring a “parsing” of the action sequence of the model, as in the “artificial fruits” studies, the second attention to the manner of the action itself. However, it appears that apes encounter more difficulty with the action level, resulting in “approximate” rather than exact imitation, resulting in arguments on whether this is really imitation rather than emulation, i.e. achieving the same goal with whatever means are most readily available (Hurley and Chater, 2005).

Using familiar motor programs in novel situations is easier than copying (and even more so than re-enacting from memory) of novel motor actions (Myowa-Yamakoshi and Matsuzawa, 1999). It has however been shown that “enculturated” apes out-perform mother-reared apes (and in some cases even children) in certain imitation tasks (Tomasello et al., 1993), suggesting a possible “identification” with humans – argued in Hobson (2004) to be a uniquely human characteristic – or more prosaically, that a lack of familiarity with human motor activities is what normally prevents apes from better imitation (Byrne, 1999). At the same time, even highly enculturated (and human-identifying) apes such as the signed language-trained gorilla Koko and the orangutan Chantek, have shown difficulties in precise imitation, and have (for example) needed to learn most of their early signs through molding rather than imitation (Patterson, 1980; Miles, 1990; Tanner et al., 2006).

Since dyadic mimesis is a form of (actual or imaginary) re-enactment, it can also be seen to be implied in domains other than imitation: in particular in the understanding of others by imagining their communicative intentions: communicative purposes. From the perspective of the “sender” this implies third-order mentality: I want you to do X by grasping my intention that you do so, cf. Grice (1989). The simplest form of this can be seen in (asymmetric) joint attention (Zlatev, 2008), in which I not only wish the addressee to attend to a particular target, but to see that I do so as well, and – in the case of ostensive behavior such as pointing – to see that I wish him to do so. In this respect apes who have not undergone enculturation (and language training) consistently fail (Tomasello et al., 1997; Tomasello, 1999). Captive apes do point, both with the whole hand and in some cases with the index finger (Leavens and Hopkins, 1999) but all documented evidence shows “imperative” pointing, where the function of the pointing gesture is to get the addressee to do something, rather than the “declarative” variety, where the function is to achieve joint attention itself (Brinck, 2003). On the other hand, apes who appear to have “grasped the point” of communicative intentions such as Koko, Kanzi and Chantek, do point declaratively, showing that the capacity for triadic mimesis is not completely beyond their competence, under admittedly very special conditions.

The evidence for the presence of the other major type of communicative gestures in apes: the iconic ones, is similar. In these gestures the expression resembles the intended referent, or at least some salient part of it – as in the so-called “metonymic gestures”, e.g. when Koko used a self-invented sign of moving her index finger along her lip for “woman”, apparently resembling putting on lipstick (Tanner et al., 2006). Importantly, of Koko’s 50 observed “invented signs” (i.e. not signs from the simplified form of ASL that she was being taught) during her first 10 years of life, 38 (76%) were judged to be based on iconicity. In the majority of these
(27), the gorilla mimed an action, while in a sizable minority (15) she “depicted the shape of an object” (Tanner et al., 2006, p. 73). Furthermore, the first were developmentally prior to the second, and similarly gestures which resembled actions preceded those which were interpreted as standing for objects.

The gorillas in the San Francisco Zoo, and above all the male Kumbie were observed longitudinally (Tanner and Byrne, 1996; Tanner et al., 2006) and some of their gestures such as the often mentioned “armswing under” (“come and touch my genitals”), could be readily interpreted as iconic with respect to the desired actions. The fact that such gestures were used with and interpreted by several different females suggests that they were not acquired through ontogenetic ritualization but rather involved an interpretation of the iconic relationship by both sender and receiver. On the other hand, these were limited almost entirely to a single individual (Kumbie) and only “meant” desired actions, and never objects. In the studies of bonobo and gorilla gestures by Pika et al. (2003, 2005), no (intended) iconic gestures were observed, though considerable group differences within the species were found, attesting for simpler mechanisms of social learning such as ritualization, which do not require understanding the sign function or communicative intentions.

In sum, at least in some contexts non-language-trained apes both point and are capable of producing and understanding gestures which resemble desired actions, but it is unclear to what extent such apes appreciate the triadic sign relationship: X stands for Y for S, and the accompanying communicative intentions. On the other hand, apes such as Koko, Kanzi and Chantek do so after prolonged exposure to sign use by human care-givers, suggesting that triadic mimesis is within the periphery of their “zone of proximal development”, and thus in the “zone of proximal evolution” of the common ancestor (Donald, 2001; Zlatev et al., 2005). This yields support for the hypothesis that triadic bodily mimesis was a crucial (“missing”) link in hominid evolution (Zlatev, 2002), as is also suggested by the attested capacity for “protolanguage” in apes, reviewed below.

3.4. Post-mimesis 1: signs and simple combinations

Evidence from four of the most successful projects involving the teaching of language to great apes – the chimpanzee (Pan troglodytes) Washoe (Fouts, 1973) the gorillas Koko and Michael (Patterson, 1980), the bonobos (Pan paniscus) Kanzi and Panbanisha (Savage-Rumbaugh et al., 1988) and the orangutan Chantek (Miles, 1990) – shows that as long as the expression side of the sign relationship is produced manually/visually, apes can acquire a form of protolanguage. As with children a precondition for the success of these projects was a cultural environment rich in intersubjective engagements and a variety of activities to stimulate communication. More specifically, the evidence is convincing that these apes:

- comprehend the 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 large active vocabulary of words/signs, according to some measurements extending 600 signs, but even according to the most conservative criteria no less than 140 signs;7
- regard the acquired signs as conventional-normative (consensual), to the point of correcting their teachers if the latter do not use these appropriately; and
- understand novel combinations of spoken or signed words.

Thus, we may agree with Miles (1999, p. 204) that all great apes “have the intelligence for a rudimentary, referential, generalizable, imitative, displaceable, symbol system”. On the other hand, while some rudiments of the following have been attested, it is also clear that the capacities of the language-taught apes are limited with respect to

- producing novel combinations of signs (in a systematic manner);
- using language for a number of different functions (speech acts): informing, requesting, asking for information, etc.; and
- using language not only for communication, but for thinking (private speech).

The latter capacities distinguish post-mimesis 1, or “protolanguage” from post-mimesis 2, which is equivalent to language, either spoken or signed. Even under the most generous interpretations, none of the enculturated apes has progressed to the attainment of this last level of the mimesis hierarchy.

3.5. Post-mimesis 2: semiotic systematization and “theory of mind”

Perhaps the most convincing evidence of producing sign combinations by the language-trained apes have involved labelling, in which a pointing gesture directed at an object, or a picture of one in a book, is accompanied with the indication of a lexigram denoting the object. Kanzi has been shown to use pointing + lexigram, as well as lexigram + pointing combinations for the sake of labelling quite frequently: there are 249 examples in his five-month corpus of two-element combinations (Greenfield and Savage-Rumbaugh, 1991). In some cases, the pointing gesture was even combined non-redundantly with a lexigram denoting an action with respect to the object or person pointed at. At the same time, it has not been demonstrated that the language-taught apes’ sign-combinations conform to any consistent principles of grammatical organization. Likewise, despite some claims to the contrary, it has not been shown that Kanzi or any of the other enculturants has passed a false-belief test (Call and Tomasello, 1999; Povinelli, 2000). There appears to be a close connection between the two abilities: language proficiency and the conceptual understanding of others’ mentality. Three different sides to language (use) combine to promote the latter. First, since language is a conventional-normative semiotic system (cf. the introductory discussion in Section 1), its mastery implies expectations on the ways others “should” behave, and the kind of knowledge they would need in order to do so. Thus, language use itself would imply habituating one to think about knowing or not knowing certain facts, e.g. the “proper” meaning of words. Second, two relevant (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 X”, then one should be able to think the corresponding thoughts. Third, as pointed out by Tomasello (1999), not just the logical 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, p. 173): disagreements, repairs/explanations and meta-discourse. Empirical evidence for the existence of a
strong connection from language mastery to the understanding of beliefs involves, e.g. the following:

- Deaf children who are not exposed to a signed language at an early age pass false-belief tasks significantly later than those who are (Peterson and Siegal, 1995).
- There are correlations between parental use of mental predicates in their child-directed speech and the children's performance in false-belief tasks (de Villiers and Pyers, 1997).
- Longitudinal studies of child development indicate that language development predicts false-belief task performance, rather than vice versa (Astoning and Jenkins, 1999).
- Training children in (non-mental predicate) sentential complement constructions significantly improves performance on false-belief tasks (Hale and Tager-Flusberg, 2003; Lohmann and Tomasello, 2003).
- Training in perspectival discourse alone (without sentential complements or mental predicates) contributes to performance on false-belief tasks (Lohmann and Tomasello, 2003).

A possible conclusion is therefore that the understanding of (false) beliefs requires post-mimesis 2, i.e. based on language, either spoken or signed, though it is less clear which aspects of language (use) are most essential. Notice that this does not contradict studies showing that the acquisition of language presupposes the so-called “theory of mind” skills (Bloom, 2000), since the latter, including joint attention and communicative intention are dyadic or triadic mimetic phenomena, which also in the present analysis are prerequisites for language (Zlatev, 2003; Gallagher and Hutto, in press).

3.6. Summary

The evidence reviewed in this section, showing both similarities and differences between ape and human capacities along the mimesis hierarchy can be interpreted as suggesting quite different developmental trajectories for apes and children. Fig. 1 presents the mimesis hierarchy as a developmental (rather than an evolutionary) progression where each new stage is not viewed as a “plateau”, but as a transition in development, leading to a change in the trajectory. It is furthermore a “layered model” in the sense of (Stern, 2000 [1985]) in which the acquisition of a new capacity does not “abolish” or “replace” earlier ones, but rather introduces a new layer of cognitive complexity, while capacities on the lower layers continue to operate “in parallel”. They are not replaced, either in evolution or in ontogeny, since they continue to be functional. Proto-mimetic skills (primary intersubjectivity) “constitute our primary access for understanding others, and they continue to do so even after we attain our more sophisticated abilities in this regard” (Gallagher and Hutto, in press). Similarly, dyadic mimesis (e.g. imitation), triadic mimesis (e.g. pointing and iconic gestures) and post-mimesis 1 (e.g. holophrases) continue to play crucial roles in human social interaction, cf. Tomasello et al. (1993), McNeill (2005) and Wray (2002). Hence, the evolutionary argument against “gesture origins of language” theories made by, e.g. Deacon (1997) that if a system develops from a previous one, it tends to replace it, does not invalidate “Russian Doll” theories such as the present one, and those of Zlatev (2003) and de Waal (2007).

In sum, the proposal is that the human and ape (and perhaps even primate) skills in proto-mimesis are not radically different, while in terms of dyadic mimesis (e.g. imitation), we clearly notice a difference in capacity, leading to a divergence of the trajectories. The second transition, to triadic mimesis, does not really materialize in the ape line, though something at least resembling it can occur in special circumstances (as in the San Francisco Zoo). At the same time, “non-typical” development in apes through human enculturation would change that so that apes could become capable of a limited degree of triadic mimesis, and with extensive training even post-mimesis 1. However, this would not place them on the same trajectory as human children, since nurture cannot completely outdo nature, manifested in differences such as those discussed in the following section, and apes will not acquire language fully, and would not be able to reason about the “states of mind” of others.

4. From proto-mimesis to language: neural bases

In this section, I summarize recent findings from neuroscience suggesting how the ascending levels of the mimesis hierarchy can be correlated with “expanding” neural circuits in the brain. I should state from the onset, however, that this will be done in a tentative manner, which is a necessary caveat, given that (a) our understanding of the neural basis of imitation, intersubjectivity, gesture and language is still at its infancy – despite the recent upsurge of activity in this field and (b) that my own understanding of this is exclusively based on second hand reports. However, these two minuses can also be interpreted positively, and thus motivate the application of the mimesis hierarchy to neuroscientific data: since (a) it can serve a useful conceptual role in clarifying certain distinctions (which are, alas, often forgotten) and (b) since I have no personal investments in the interpretations given to the findings of any particular research group. I will similarly interpret evidence broadly to include single-neuron activation measurements in monkeys, clinical evidence from patients, transcranial magnetic stimulation (TMS), various forms of “brain imaging” and even computational modeling. A “missing link” in the evidence that will for obvious practical reasons remain absent for some time are non-invasive studies of the living brain of great apes (and for obvious ethical reasons, invasive ones will/should not be available either).

The existence of perception-action cycles in animal and human brains has been known at least for a century (Berthoz, 2000), but has been – with some notable exceptions, e.g. (Arbib, 1972) – bypassed, probably due to the “modular” approach focusing on separate “functions” and brain areas, that has dominated both traditional neurology and modern cognitive neuroscience. If only for going against this tendency, the discovery of “mirror neurons”

---

Fig. 1. A schematic representation of the divergent developmental trajectories of (typically developing) children and apes in terms of the stages of the Mimesis Hierarchy model (the two levels of Post-mimesis are not distinguished).

---

8 However, recent studies involving PET imaging of monkeys (Gil-da-Costa et al., 2006) may point the way to possible forays into ape social neuroscience. I thank Michael Arbib for this reference.
in area F5 of the macaque premotor cortex (di Pellegrino et al., 1992; Rizzolatti et al., 1996), firing during executed as well as observed goal-directed actions deserves at least some of its wide acclaim. Still, it is amazing how these cells have been proposed as a solution to just about every mystery in the human mind: from empathy to imitation, mind-reading, language (evolution), autism, and sexual preferences! It is not surprising that their role has been regarded as much over-rated by some researchers in the field (Preson and de Waal, 2002; Donald, 2005; Csibra, 2007).

Additional studies have shown neurons with similar properties in the parietal cortex (area PF) of the monkey brain (Fogassi et al., 1998), “audiovisual” mirror neurons, responding to either the observation or the sound of the action (or both) in F5 (Kohler et al., 2002) and a multitude of studies involving TMS, e.g. Fadiga et al. (1995) and brain imaging, e.g. Buccino et al. (2001) have suggested the presence of a homologous mirror neuron system (or systems) in human brains, and additionally found mirror properties for emotions in the anterior insula (Wicker et al., 2003) and (secondary) sensorimotor cortex (Adolphs, 2003).

Such findings are impressive, but despite suggestions to the contrary, a “unifying view of the basis of social cognition” has been lacking. When, for example, Gallee, Keyserz and Rizzolati write “When only the cortical centers, decoupled from their peripheral effects, are active, the observed actions or emotions are ‘simulated’ and thereby understood” (Gallee et al., 2004, p. 400), this is based on the assumption that neuron firing in itself possesses “representational content” (Gallee, 2005; Gallee and Lakoff, 2005) which is doubtful: it is the experimenters who attribute this “content” on the basis of their observation of the temporal co-occurrence, i.e. a form of “indexicality” (Sonesson, 2007) between events in the world and neural patterns, not the animal, and not the (human) subject. The fact that mirror neurons fire during either observations/sounds on the one hand and executions of actions on the other, does not make them more representational, than say, neurons in the visual cortex responding the particular aspects of the observed scene. Admittedly, the action/observation co-relation is an indication for some kind of “parity” of self- and other-executed actions, but to be a (true) representation there needs to be a subject distinguishing being expression and content, and needless to say no homunculus in the brain playing this role has yet been (convincingly) shown.

The second reason for skepticism has been in the much-too-quick jump from actually attested “mirror neurons” in monkeys and the human studies using indirect evidence to the conclusion that human brains have a “mirror neuron system”. As observed, monkeys have a mirror neuron system yet do not imitate (Visalberghi and Fargaszy, 2002), while for example Iacoboni’s (otherwise very suggestive) “minimal human architecture for imitation” (Iacoboni, 2005) involves only human homologues of the monkey’s mirror system. It has been argued that the human capacity for imitation was built “atop” the monkey’s system for action recognition (Rizzolatti and Arbib, 1998; Arbib, 2005), but the explanation of what this actually means have been rather vague, and mostly quantitative, while at the behavioral level difference between monkeys (and possibly also apes, as reviewed in Section 3) and humans with respect to imitation appear to be qualitative (Decety and Chaminade, 2005).

Despite these words of warning, I will in the following adopt the general strategy of Rizzolatti and Arbib (1998) and Arbib (2005) of “scaling up” from what we know about social resonance in monkeys, through (proto-) imitation in apes, to true/full imitation, representation and language in humans along the mimesis hierarchy presented in the previous two sections. By quoting empirical findings, and filling in with some hypotheses (i.e. speculation), I will suggest transitions which are more qualitative than quantitative. The key point is not to over-interpret the role of mirror neurons but to see them as part of the puzzle in bridging the wide differences between monkey and human brains – and minds. It seems that some progress has been made in the field when instead of thinking about single neurons or even areas, researchers have begun to meaningfully connect these in circuits, and applied principles from dynamic systems theory and neural net modeling to the understanding of the role of these circuits (Arbib, 2005; Iacoboni, 2005; Hurley, 2005).

4.1. Proto-mimesis: dexteral manual control and social resonance

Despite a number of disagreements between researchers in social neuroscience, there seems to be convergence on the existence of a frontal-parietal-temporal system in the monkey’s brain whose initial function was increased manual dexterity, and secondarily “resonance” with the actions of others, providing a kind of implicit recognition or parity, without involving any representational thought. The key components are the now famous area F5, segregated in “canonical” (responding to self-actions and observed object-affordances) and “mirror” (responding to self-actions and congruent observed actions) sub-areas, the superior temporal sulcus (STS), responding to the observation (though not execution) of acts of biological motion, and connected to both of these reciprocally: area PF in the rostral part of the inferior parietal lobe, which is highly “functionally heterogeneous” (Rizzolatti, 2005, p. 59), responding to both somatosensory and visual stimuli, with a subset of the neurons discharging also during motor activity, so-called “PF mirror neurons” (Fogassi et al., 1998).

The important features of the F5-PF-STS circuit, connected also strongly with the prefrontal cortex (PFC), is that it forms an “observation/execution matching system” (Arbib, 2005), which in combination with exploratory (manual) sensorimotor activity, can “self-train”, using the efferent copy, or “forward model” (Kawato, 1999) from the canonical neurons as a means to form more skillful goal-directed hand movements and in the process to recruit mirror neurons responding to the combination of actions and objects: “As a result of this training, the appropriate mirror neurons come to fire in response to viewing the appropriate trajectories even when the trajectory is not accompanied by F5 canonical firing” (Arbib, 2005, p. 112). Since the trained neurons respond to the relation between the hand trajectory and the object, they become neutral with respect to whether the hand belongs to the “self” or the “other”, thereby yielding their “mirror” properties. Such a developmental process can account for the learning of an open-ended number of object-directed hand movements, thereby avoiding the highly implausible scenario of a pre-specified set of neurons (assemblies) for particular actions. In the case of “mouth mirror neurons” (Ferrari et al., 2003), some of these may be pre-tuned innately (i.e. prior to birth), thus possibly accounting for neonatal mirroring, given that these neurons respond to lip- and tongue protrusion, which are the imitated actions most robustly established in new-born babies, chimpanzees (Myowa-Yamakoshi et al., 2004), and macaques (Ferrari et al., 2006).

To emphasize again, such a system yields a basis firstly for flexible motor control, and secondly a means for the automatic recognition of certain actions and communicative signals in con-specics. The “high-resonance” aspects of it, focusing on the goals of actions, would serve the basis for emulation (Hurley, 2005; Csibra, 2007), while the “low-resonance” aspects, could possibly be a basis for automatic processes of contagion and mimicry, as suggested by Rizzolatti (2005). In other words, this could be the neural basis for proto-mimesis (cf. Table 1 in Section 2), based on a “mapping” between exteroception (STS) and proprioception (PF), as well as motor control (F5). This system is also densely interconnected with a sector of the insula, which on its part connects with the amygdala (Augustine, 1998), thus possibly accounting for the mirror system’s
role in the communication of affect, i.e. a basic form of empathy (Iacoboni, 2005).

What this does not yield is anything more complex such as imitation or perspective-taking, and definitely not representation (as defined in this chapter). The reasons for this are that there is yet no (conscious) differentiation – between self and other, and between expression and content. While there may be volition in the case of praxis, there is yet none in the case of imitative or representational activity, where one's own body movements need to be carefully controlled to match those observed. To the extent that we can talk about "embodied simulation" (Gallese, 2005) on this level, it is solely of the “sub-personal” kind, of the type described by, e.g. Berthoz (2000).

4. Dyadic mimesis: imitation and imagination

Iacoboni (2005) states that the perisylvian cortex, in which the F5-PF-STS circuit is located in the part of the brain that has proportionally increased in volume the most when one compares the brains of chimpanzees and humans, while Deacon (1997) estimates that it is rather the PFC. Given the strong connections between the two, it is likely to be a combination of both (Decety and Chaminade, 2005). As mentioned, we can not yet study the brain "in action" in apes, so let us turn to the human data, attempting to infer on the basis of indirect evidence what the "intermediary" ape level (of course, used as an approximate for the common ancestor) of the neural basis for dyadic mimesis could be like.

The human homologue of F5 has been argued to be the pars opercularis of the inferior frontal gyrus (BA 44) (Rizzolatti and Arbib, 1998), which has been shown to be active during both observation and imitation of hand actions (Iacoboni, 2005). The pars opercularis is differentiated in a dorsal part (with "mirror" properties) and a ventral part which is active only during imitation but not observation. Iacoboni (2005) suggests that this ventral part may be the source of the efferent copy of the motor action projected to the parietal cortex. The latter itself differs considerably from that of the monkey (Decety and Chaminade, 2005). Both the superior and the inferior human parietal cortex is active in observation, action and imitation tasks, suggesting projections to both from inferior frontal cortex, while in the monkey brain, F5 projects only to the inferior parietal cortex (Rizzolatti, 2005). The human inferior parietal cortex, especially in the right hemisphere, is crucially involved in tasks requiring imagination ("consciously representing actions", Decety and Chaminade, 2005, p. 127), and not just sub-personal "simulations". Importantly, when subjects are asked to imagine themselves performing an action, it is the left inferior parietal lobule (together with left somatosensory cortex), that is activated, while when imagining someone else performing an action, it is the right IPL that is mostly activated, together with posterior cingulate and frontopolar cortex. On the other hand, the parietal cortex is not limited to imagination tasks, but is also active in action execution. One factor distinguishing the two (action vs. imagination) is activation in posterior cerebellum which together with the PFC possibly serves to inhibit action during imagination tasks (Lotze et al., 1999; Decety and Chaminade, 2005).

Lateralization seems to be important for imitation, as shown by the studies of Decety and colleagues mentioned above. Another important area is the superior temporal gyrus (STG), which also plays somewhat different roles in the two hemispheres, as suggested by Decety and Chaminade (2005, p. 135) "the right STG is involved in genuine visual analysis of another's actions, while the left region is concerned with analysis of the other's actions in relation to the intention of the self". Inherited from our ancestors in the network is the STS, which is more highly active during imitation than in observation, suggesting that the efferent copy from inferior frontal cortex (via parietal cortex) is used not only for "simulating" action, but for matching observed action and imitated action (Iacoboni, 2005). Finally, medial PFC and STS are also considered crucial for the "mentalizing" system in humans (Castelli et al., 2002).

The purpose of the review in the above paragraphs was to suggest how extended the human "mirror neuron system" is in comparison to its predecessor in the monkey/human common ancestor (for which present day monkeys is our best approximation). Furthermore, none of the neural structures described, and the functions they support (imitation, imagination, and perspective-taking) seem to involve language in a crucial way. Suggestions to the contrary, e.g. concerning "unconscious verbalization" (Heyes, 2001) leave empirical findings unexplained (Rizzolatti, 2005; Iacoboni, 2005). The natural conclusion is that much (if not all) of this system evolved prior to language. The difficult questions are "which parts" and "when", and given the lack of comparative evidence from non-human apes, we can only formulate best guesses. The ones I propose are the following: given the recent positive evidence on chimpanzee understanding of others' perspectives and possibly even "mental time travel" as well as for "simple" imitation (cf. Section 3), together with the previously established results of mirror self-recognition, it is likely that large parts of this extended system are present in apes, at least in rudimentary form. On the other hand, given the lower degree of lateralization in the chimpanzee brain, with 50:50 lateralization with respect to handedness (Marchant and McGrew, 1996), and the prefrontal and perisylvian cortex expansion in Homo, combined with the behavioral evidence of a limited capacity for imitation in apes, the chimpanzee version of the system must be also limited compared to the one sketched out above. For example, it is possible that the chimp F5/pars opercularis IFG homologue is not differentiated in a ventral and dorsal part, leading to a "lower level" efferent copy which would account for difficulties with action imitation especially when the action is not performed in relation to objects.

On the other hand, by linking imitation and perspective-taking to lateralization, as done most prominently by Decety and colleagues, we can interpret the fossil evidence of lateralization for handedness in Homo ergaster (cf. Johansson, 2005) as indicating enhanced capacities for dyadic mimesis. Concerning the debate on whether intersubjectivity or imitation presupposes the other (Hurley and Chater, 2005), my answer would be: neither. Social mimicry and social resonance (as aspects of proto-mimesis) can be said to presuppose both, but the capacity for developed imitation and perspective-taking could have come in tandem in evolution, as they appear to co-emerge and mutually enhance one another in ontogeny (Piaget, 1945; Zlatev, 2008). Notice that both imply an understanding, a "concept" (cf. Whiten, 2000) of iconic representation ("this is like that, but also different from it"), above all in the visual-manual domain. From here to the use of actions for intentional communication is not such a large step.

4.3. Triadic mimesis: iconic bodily signs

A well-established fact is that F5 mirror neurons in the macaque do not respond to "intransitive", non-object related manual actions, while they do so for "transitive" ones, whether the object is visible or was only recently visible, and in the present occluded (Umlită et al., 2001). Thus, they would not respond to pantomimed actions. Galles (2005) refers to mirror neurons in the ventral part of F5 responding to self or other lip- and tongue-protrusion and lip-smacking (Ferrari et al., 2003) as mirror neurons for "intransitive" actions, but as suggested in Section 3, these appear to be...
either Innately specified or quickly attuned to support proto-mimetic interactions, and can not serve as an open-ended system of communication, and even less so for expression-content differentiation. On the other hand, the neural system supporting dyadic mimesis would necessarily need to allow for the imitation of intransitive actions — even for their own sake, and not in relation to a goal, as in the case of do-as-I-do imitation. This copying (or mentally re-enacting) of the “means” is the first requirement in order make the step to pantomime, which is based on an iconic relationship between the motor act and its semiotic — rather than praxic/functional — meaning. The second is to be able to make (for the mimic) and to perceive (for the audience) the iconic relationship explicit, i.e. to differentiate between the expression and the content. Since iconic gestures do not have a praxic meaning, i.e. a functional goal, we are led to interpret them as meaningful in relation to what they stand for, i.e. represent, properly speaking.

What are the neural bases for pantomime and (iconic) gesture? There seems to be a relative paucity of evidence for this, which is surprising given the large amount of energy devoted to, on the one hand, pre-semiotic capacities such as empathy and imitation, and on the other hand, language: a “missing link”. McNeill has some suggestions (McNeill, 2005), but since his theoretical model ties language and gesture by definition, the two are difficult to tease apart even in the presentation of the data. In the chapter on “Neurogesture”, however, he highlights the role of the right hemisphere (in right-hand subjects). Patients with RH-damage are impaired not only in terms of narrative coherence (Garnder et al., 1983) but also produce twice as few gestures. In the two “split brain” patients reported, McNeill observed an interesting decoupling between speech and gesture: one patient seemed to rely mostly on his left hemisphere, and produced few, small and non-typical gestures, except when he became silent, and gesticulated profusely, which McNeill calls metaphorically a possible “right hemisphere coup”. The other patient spoke less, but “had gestures of greater iconicity” (McNeill, 2005, p. 229). The interpretation suggested is that the “imagery” (i.e. the content of the gesture) is evoked within the right hemisphere, while in normal subjects it is “orchestrated” by Broca’s area (BA 44 and 45) and in the case in which it is accompanied by speech, integrated with it. But given a stage in evolution where speech was still not recruited for intentional communication, the present day segregation of BA 44 and 45 would hardly have been present (or would have had a different function from that in modern humans, see next section). Thus, the neural substrate for iconic gestural communication (including pointing) would have been the same system for dyadic mimesis that was described in Section 4.2, involving the frontal-parietal-temporal circuit for action-perception matching, with a left–right hemisphere division of labour. The only difference is that the frequent employment of (above all) hand movements as representations of non-present actions, objects or events, would make this “division of labour”, i.e. lateralization, more pronounced. While the RH could be more involved in representing the (potentially conscious) imagery, i.e. the content of the gestures, the LF could specialize in their expression. Indeed, Choi et al. (2001) found activations of premotor and parietal areas when performing pantomimic gestures compared to meaningless finger movements, which were lateralized to the dominant hemisphere independent of the hand used in gesturing. Thus, my proposal is that hemispheric laterality is not only essential for full imitation and intersubjectivity, as suggested by Decety and Chaminade (2005), but for the separation of the “signified” and the “signifier”, i.e. for the emergence of the sign function (Piaget, 1945). And given its evolutionary history, this would have been through triadic mimesis, i.e. gesture and pantomime.

Notice, however, that triadic mimesis does not equal a full “gesture language” or “protosign”, and is thus immune to the criticism delivered to the proposals of Corballis (2002) and Arbib (2005) concerning the “indirectness” of the evolutionary path to speech through gesture (Seyfarth, 2005), the presence of neural specializations and even “mirror neurons” for speech processing (Wilson et al., 2004) and the non-“supplanting” but rather integrating of speech and gesture (McNeill et al., 2005), see Section 5. On the other hand, it is more than simply “pantomime” since in combining a pointing and an iconic gesture, or two iconic gestures, a “proto-proto-language” consisting of simple predications would have been possible. These however need not have been conventionalized, but would have to be re-negotiated on a case-by-case basis.10 The pressure to “disambiguate” them would have been one of the factors to provoke the recruitment of vocalizations, and eventually speech.

4.4. Post-mimesis 1: protolanguage

The term “Broca’s area” is still widely used in textbooks, but there is now reasonable agreement that it should be both anatomically and functionally divided in the pars opercularis (BA 44), the human homologue of macaque area F5, and the pars triangularis (BA 45) of the inferior frontal gyrus. Whereas the second consist of “higher-order heteromodal association cortex suited to complex cross-modal associations typical of higher-order linguistic functions (syntax, lexical semantics)” (Barrett et al., 2005, p. 125), the first has been found in an fMRI study to be more specifically specialized for articulatory and motor speech functions (Paulesu et al., 1997). The (partial) dissociation of BA 44 and 45, points to a specific human adaptation for speech, and since the function of speech could have been initially to “disambiguate”, and thus stabilize and conventionalize meanings, this is an indication that this is part of the neural substrate for post-mimesis rather than triadic mimesis. On the other hand, since BA 44 is also activated during the observation and imitation of hand gestures, it cannot be so strictly “specialized”, thus suggesting an evolutionary “from hand to mouth” scenario, in which BA 44 evolved these functions from initial control of bodily mimesis rather than from the primate system for vocal communication. The latter cannot be excluded, given the finding of “mouth mirror neurons” (Ferrari et al., 2003) and the “audiovisual mirror neurons” (Kohler et al., 2002), but the proposal for a direct evolution from this to “phonetics” leaves unexplained the ability to share the meaning of expressions through the primary iconicity (Sonesson, 2007) of gestures, and the integration of speech and gesture (McNeill, 2005), as well as the non-speech functions of BA 44. Another study (Mesiter et al., 2003) using TMS offers additional support for left-hemisphere recruitment of speech functions by the “hand motor area”: during reading two syllable words, the latter was activated, but not the leg motor area, while in making non-verbal oral movements, motor cortex was slightly activated bilaterally.

Other areas outside “Broca’s area” have been found to be involved in speech processing. A recent study (Wilson et al., 2004) gives support for the “motor theory of speech perception” (Lieberman et al., 1967) in finding (non-lateralized) areas in the superior portion of the ventral premotor cortex (PMv), which were active during both perception and production of meaningless syllables (BA 4a and 6), while an adjacent more ventral area (BA 6) was activated during the production of the same syllables. Significantly, the activity in all these areas was “much greater for speech than for non-speech sounds” (Lieberman et al., 1967, p. 702). Furthermore, strong activations were also found in the superior temporal gyrus.

10The closest illustration of this that I can think of is the early gestural communication of deaf children to non-signing parents, developing a “homesign” (Goldin-Meadow, 2003). For an illustrative video see: http://goldin-meadow-lab.uchicago.edu/research_state.html.
(STG) during listening, as well as activation in BA 44, for 9 of the 10 subjects (which the authors call “less consistent” since for five subjects the activations were in the left hemisphere, and for four in the right hemisphere, see Supplements to Wilson et al., 2004). This last fact is significant, since it suggests an extension of the perception/action mapping network supposedly underlying bodily mimesis, rather than a separate adaptation. In the terms of Donald (2001, 2005), this is an extension that could support “voco-mimesis”. Thus, again, the model presented in this paper suggests how the different findings and theoretical perspectives can be unified in a single picture, rather than being posed against one another.

The other classical “language area”, Wernicke’s in the superior temporal gyrus, seems to have received much less recent attention. Arbib (2005) suggests that it corresponds to the posterior part of BA 22 (area Tpt, temporo-parietal), “but also (in whole or in part) to areas of the human cortex that corresponds to macaque PF” (Arbib, 2005, p. 121), thus implying that it is an integrated part of the human mirror neuron system specializing for “recognizing protosign and protospeech to support a language-ready brain that is capable of learning signed languages as readily as spoken languages” (Arbib, 2005, p. 121). This sounds plausible, but is underdeveloped. Unless the human inferior parietal lobule (IPL) is to be included in Arbib’s “PF extended”, what evidence is there that it is included in the execution/recognition network, apart from the well-known projections from Wernicke’s to Broca’s area via the arcuate fascicle? Clearly this is an area for further research, and its relative neglect may be due to a too narrowly specified focus on “mirror neurons”. In any case, since Wernicke’s area belongs to the perisylvian cortex, strongly expanded in the human brain (see Section 4.2), it has without doubt been adapted to new functions from its predecessor in the ape-human ancestor, with the recognition (and generation) of both speech and gestures being a likely candidate.

4.5. Post-mimesis 2: language

If post-mimesis 1 corresponds to simple “protolanguage” that involves gestures as well as speech, what would be necessary to extend this to a modern kind of language with grammatical structure, morphology, recursion, etc? Here I find Arbib’s (2005) proposal that this could have been (mostly) “post-biological”, i.e. as resulting from cultural, rather than biological evolution during the past 100,000 years quite plausible and the most parsimonious alternative. The difficulties of Arbib’s specific scenario according to which “protosign” and “protospeech” were exclusively holophrastic, and analytic language developed through “fractionization”, could be resolved if it is remembered that triadic mimesis, and thus post-mimesis 1 was not completely “structureless”, but could include gestural-vocal utterances that are composed of at least two expressions, thus forming predications (e.g. “tiger” + point). The multimodal communication of young children in the “one word stage” (Goldin-Meadow and Butcher, 2003), and possibly the combinations of lexigrams and pointing performed by Kanzi (Greenfield and Savage-Rumbaugh, 1991) indicate a more plausible starting point for the cultural evolution of grammar than (only) holophrases. All attested cases of grammaticalization – language change in which “open class” lexical items, e.g. going to receive grammatical meanings, e.g. Future tense – can be ultimately traced back to “action words” and “object words” (Heine and Kuteva, 2002), and these are present in the two-part division of predications available in post-mimesis 1.

On the other hand, exaptations (pre-adaptations) for grammar can be found not only in the hierarchical structure of motor praxis, e.g. gorilla nettle eating (Byrne, 1999), but even more so in the ape’s attested capacities for the imitations of sequences of actions, which also seem to involve a degree of hierarchical structure (Whiten et al., 2004). In the evolution of dyadic mimesis, apart from the differentiation between self and other, and means and goals (emphasized in Section 3.2) the neural basis for such imitative acts would have been further extended to include (minimally) the basal ganglia, pre-supplementary motor area (pre-SMA), and cerebellum. In this sense, with the establishment of post-mimesis 1, the human brain would have been “language ready” (Arbib, 2005), and the transition to post-mimesis 2 (mostly) a matter of socio-cultural evolution. Having said this, the most parsimonious hypotheses are not always the correct ones, and it is possible that language-related (though not necessarily language-specific) circuitry enhancing the learning of sequences and hierarchical structures were selected for in the transition to “full” language as well. Somewhat surprisingly, the relatively well-preserved grammatical capabilities of high-functioning children with autism suggest such a secondary “reserve” route to language, i.e. not primarily via bodily mimesis (Kjelgaard and Tager-Flusberg, 2001; Järvinen-Pasley and Heaton, in press; Zlatev, 2008).

4.6. Summary

This section presented a summary of findings related to the (possible) neural underpinnings of the different stages of the mimesis hierarchy. As pointed out at the beginning, they are all tentative, and some are speculative. But the interpretations are (I believe) coherent, and can possibly help clarify and even guide ongoing research in social neuroscience. Instead of looking for either “theory of mind modules” or content ourselves with cataloguing ever increasing areas with “mirror neurons”, I have suggested how it is possible to view the neural bases of bodily mimesis as an interconnected system that links many parts of the brain, and that has gradually expanded in terms of areas and functions throughout evolution. At the basic proto-mimetic level there is a frontal-parietal-temporal-amygdal perception/action matching system that is used for motor control and social resonance. With the recruitment of new areas and hemispheric lateralization, the system became extended to make (true) imitation and imaginative re-enactment possible, at least the rudiments of which seems to be available to apes. The next step, triadic mimesis, involves utilizing the system for intentional communication on the basis of, above all, iconic gesturing. Post-mimesis 1 involves the establishment of conventional (normative) signs with the help of speech, and eventually grammatical language, post-mimesis 2. Throughout this process, motoric structure became extended from praxis, through imitation to utterance, including both gesture and speech (Kendon, 2004), and in the process conventionalized and systematized. In adult human beings, this system involves areas in just about every part of the brain, and in both hemispheres. Some disruptions of it (including such large ones as removing the right hemisphere in early childhood) may not have devastating effects, while such that impair the system at its “core” and at an early age are likely to have lasting consequences. One possibility is that autism results from such impairment (Williams et al., 2001; Zlatev, 2008).

5. Evolutionary implications

The previous two sections have implicit in them the bases for a particular scenario for the evolution of language, which I here will spell out quite briefly, for lack of space. In quite a few ways it is similar to those of Donald (1991), Corballis (2002) and Arbib (2005), but as pointed out in Section 4.4, it provides answers to some of the objections directed at these theories. The capacity for proto-mimesis is largely continuous between monkeys, apes and human beings, underlying our common heritage as extremely social and dextral species (cf. Section 3.1). With
bipedalism, the human ancestors parted company with the apes, and in the context of terrestrial group life, and possibly increased infant dependency (Johansson et al., 2006) developed these capabilities even more. In the process, they evolved a unique capacity for dyadic mimesis, used for both imitation and identification with others, both of which are essential for the transition from social to “cultural” learning (Tomasello et al., 1993). For this purpose, considerable neural adaptations such as those discussed in Section 4.2, would have been necessary.

The key development toward language was, however, triadic mimesis involving declarative pointing and iconic gestures (pantomime), serving as the main means of communication of Homo ergaster/erectus. Since triadic mimesis constitutes not an “arbitrary”, but a highly motivated and transparent semiotic system, based on the iconicity (pantomime) and indexicality (pointing) of gestures, and their combination, this avoids the objection that the primate or even an extended “mirror neuron system” can at most support the understanding and sharing of actions but not of signs (Hurford, 2004).

At some point in evolution “vocal gestures” became increasingly recruited to supplement, rather than “replace” gestural communication. This would have at first been a gradual process, but with the emergence of Homo sapiens some 200,000 years ago, it would have been firmly established, and with it the emergence of an integrated gestural-vocal protolanguage (McNeill, 2005), which constitutes the transition to post-mimesis 1. Apart from all the ecological benefits often quoted (such as communication in the dark), the use of speech as the main means of communication in hearing individuals would have been enhanced by the greater arbitrariness of the expression side of the sign relation (Corballis, 2002; Zlatev, 2002), further distancing communication from action. This on its part would have promoted the need to rely on semiotic conventions and norms of communication, thereby further enhancing our capacity for intersubjectivity (cf. Section 3.5).

Notice, that this scenario does not imply that at any stage of evolution there was a purely gestural language, i.e. a conventional-normative semiotic system. There is no need to suppose a transition from pantomime to “protosign” consisting of “arbitrary gestures” (Arbib, 2005), since (a) that would have led to an evolutionary drift towards signed rather than spoken languages and (b) “modern” gestures are clearly not arbitrary (save for “emblems” which are a recent, and rather peripheral, cultural invention). Rather, the proposed transition was one directly from triadic mimesis to a gestural-vocal protolanguage, which is also the “compromise” suggested by McNeill et al. (2005). The long-lasting controversy on whether protolanguage was mainly gestural or oral, whether it was holophrastic or analytic, cf. (Cangelosi et al., 2006), can thus be resolved by agreeing that it was both: gestural and vocal, and not completely holophrastic, while not as complex as modern languages. Therefore, neural correlates for speech in humans such as those described in Section 4.4 are to be expected.

Similar to Arbib (2005) I suggest that the further transition to post-mimesis 2 was largely cultural. However, in the proposed scenario no transition from a holophrastic “protosign” to “protoplasm” is necessary, but rather one between prediagnostic (e.g. pointing-to-referent + miming-an-action) non-arbitrary and not fully conventional triadic mimesis to similarly prediagnostic, but multimodal and conventional (object-sign + action-sign) protolanguage. From here, we can more clearly see the basic ingredients for a prolonged process of cultural evolution which would lead to “modern” grammatical and systematic languages. The hierarchical aspects of language are not such an evolutionary novelty as some linguists tend to claim, e.g. Hauser et al. (2002), since they could have been largely exapted from the hierarchical structure of action and imitation, which is in line with the neuroscientific evidence concerning the function of “Broca’s area”. This tendency towards systemativity would have been further enhanced by the use of a conventional, rather than an iconic, semiotic system (Deacon, 2003). What was necessary for grammatical language to evolve was first a prolonged process of brain-culture co-evolution, and then a shorter period (evolutionary speaking) in which language as a semiotic system evolved in a process of (mainly) cultural evolution. This is, in very brief, my proposed account of the path from proto-mimesis to language.

6. Summary and conclusions

In this article, I have argued that the concept of bodily mimesis and the mimesis hierarchy can be both supported by and help make sense of evidence from primatology and neuroscience and that it suggests a particular evolutionary scenario: from proto-mimesis, which is largely shared by primates, to mimesis proper for which humans are uniquely adapted, though not as categorically as for language, to language itself. Since language according to this model is essentially an advanced form of post-mimesis, it is to be expected for it to share the neural underpinnings of its predecessors, including those for action (understanding), imitation and gesture. At the same time, it is important not to try to reduce it to these structures, since each new stage in the mimesis hierarchy has “added” something new, also in neural terms.

We can now return to the two questions posed in the introduction. First, how can we reconcile the conception of language as a conventional-normative semiotic system with a perception/action-based account of at least part of its structure and meaning? This is a difficult question, since symbols are social and representational, while perception and action is individual and non-representational. Therefore, language cannot be directly “grounded” in perception and action, nor in mechanisms supporting motor control and social resonance such as mirror neuron networks, despite claims to the contrary (Gallese and Lakoff, 2005). However, if bodily mimesis mediated between sensorimotor cognition and language (Ikegami and Zlatev, 2007), it can provide the “missing link” between them, since it offers the basis for forming iconic, bodily representations, which can furthermore be the bases for establishing conventions since they are based on overtly observable and imitable and covertly experienced structures which can be shared between the members of a community. Assuming that language arose from such structures, one would expect exactly the kind findings of “embodied meaning” and gesture-language integration that recent research has established, given that language did not substitute for, but build on top of them. In claiming that language is “grounded” in bodily mimesis (Zlatev, 2005), however, one should also emphasize how it is different from it, since language is social, normative and systematic in a way which transcends mimesis, and even more so actions.

The second and perhaps even more difficult question was: why did language need to arise this way? I wish to finish with the bold conjecture that, since (a) language is a social-normative semiotic system (rather than a “computational device”, an “instinct” etc.) and (b) it evolved in a highly social and dextral primate, this was the most natural way for it to arise. Bodily mimesis offered (at least) the potentials of intersubjectivity, primary iconicity and (proto-prediagnostic structure that were essential for the transition from the sensory-motor cognition of our primate ancestors to the use of signs in our dealings which each other and with the world – a practice which eventually made us into human beings.

Acknowledgements

I wish to thank Göran Sonesson, Mats Andrén, Michael Arbib, Tatjana Nazir and an anonymous reviewer for helpful comments.
on a previous version of this text. This work was supported by the EU-project Stages in the Evolution and Development of Sign (SED-SU), cf. http://project.sol.lu.se/edsu.

References


