Journal of Physiology - Paris 102 (2008) 4–20

Contents lists available at ScienceDirect

Journal of Physiology - Paris

journal homepage: www.elsevier.com/locate/jphysparis

From grasp to language: Embodied concepts and the challenge of abstraction

Michael A. Arbib *

Computer Science, Neuroscience and USC Brain Project, University of Southern California, Los Angeles, CA 90089-2520, USA

ARTICLE INFO

Keywords:
Grasp
Language
Embodied concepts
Abstraction
Mirror neurons
Gestural origins

ABSTRACT

The discovery of mirror neurons in the macaque monkey and the discovery of a homologous "mirror system for grasping" in Broca's area in the human brain has revived the gestural origins theory of the evolution of the human capability for language, enriching it with the suggestion that mirror neurons provide the neurological core for this evolution. However, this notion of "mirror neuron support for the transition from grasp to language" has been worked out in very different ways in the Mirror System Hypothesis model [Arbib, M.A., 2005a. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics (with commentaries and author's response). Behavioral and Brain Sciences 28, 105–167; Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. Trends in Neuroscience 21(5), 188–194] and the Embodied Concept model [Gallese, V., Lakoff, G., 2005. The brain's concepts: the role of the sensory-motor system in reason and language. Cognitive Neuropsychology 22, 455–479]. The present paper provides a critique of the latter to enrich analysis of the former, developing the role of schema theory [Arbib, M.A., 1981. Perceptual structures and distributed motor control. In: Brooks, V.B. (Ed.), Handbook of Physiology – The Nervous System II. Motor Control. American Physiological Society, pp. 1449–1480].

1. The mirror system hypothesis, briefly recalled

Any normal child reared in a human society will acquire language. Some argue that this is because Universal Grammar – the set of basic structures of the grammars of all possible human languages – is innate, so that the child need simply hear a few sentences to "set the parameter" for each key principle of the grammar of her first language (Baker, 2001; Chomsky and Lasnik, 1993). Others have argued that the modern child receives rich language stimuli within social interactions and needs no innate grammar to acquire the particular sounds (phonology) of the language, and then masters an ever increasing stock of words as well as constructions that arrange words to compound novel meanings. In either case, there is something unique about the human brain which makes it *language-ready,* in the sense that a human child can learn language while infants of other species cannot. We use a comparison of human brains with those of macaque monkeys to introduce one account of how biological evolution yielded the human language-ready brain (see also (Arbib and Bota, 2003; Deacon, 2007)).

The system of the macaque brain for visuomotor control of grasping has its premotor outpost in an area called F5 which contains a set of neurons, *mirror neurons*, such that each one is active not only when the monkey executes a specific grasp but also when the monkey observes a human or other monkey execute a more-or-less similar grasp (Rizzolatti et al., 1996). Thus macaque F5 contains a *mirror system for grasping* which employs a similar neural code for *executed* and *observed* manual actions. It is important to note that in addition, F5 contains the control of *canonical neurons* which are active for execution of grasps but not for observation of the grasps of others, and other classes of neurons as well. Canonical and mirror neurons are anatomically segregated to distinct subregions F5ab and F5c, respectively, of area F5.

The region of the human brain homologous to macaque F5 is thought to be Brodmann area 44, part of Broca's area, traditionally thought of as a speech area, but which has been shown by brain imaging studies to be active also when humans either execute or observe grasps. It is posited that the mirror system for grasping was also present in the common ancestor of humans and monkeys (perhaps 20 million years ago) and that of humans and chimpanzees (perhaps 5 million years ago). Moreover, the mirror neuron property resonates with the *parity requirement* for language – that what counts for the speaker must count approximately the same for the hearer. In addition, normal face-to-face speech involves manual and facial as well as vocal gestures, while signed languages are fully developed human languages. These findings ground "The Mirror System Hypothesis" (Arbib and Rizzolatti, 1997; Rizzolatti and Arbib, 1998): *The* parity requirement for language in humans is met because Broca's area evolved atop the mirror system for grasping which provides the capacity to generate and recognize a set of actions.

---

* Tel.: +1 213 740 9220.
E-mail address: arbib@pollux.usc.edu

0928-4257/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jphysparis.2008.03.001

In putting parity at stage center in this account, we adhere to the view that the primary function of language is communication. Others have espoused the alternative view that language evolution could have obeyed an adaptive pressure for developing higher cognitive abilities and that verbal communication would be a secondary benefit. I have two comments. (i) Language is a shared medium, and thus parity is essential to it. No matter how useful a word may be as a tool for cognition, we must learn the word in the first place; and we must then engage in numerous conversations if, in concert with our own thoughts, we are to enrich our understanding of any associated concept and our ability to make fruitful use of it. (ii) Having said this, I readily admit, as is clear from the preceding, that language is a powerful tool for thought (though much thought is non-verbal). Thus, while I believe that parity was the key to getting language (or, more strictly, protolanguage – see below) "off the ground", both the external social uses of language and the internal cognitive uses of language could have provided powerful and varied adaptive pressures for further evolution of such capacities as anticipation, working memory, and autobiographic memory as language enriched both our ability to plan ahead, explicitly considering counter-factual possibilities, and mulling over past experience to extract general lessons. Indeed, where we lay stress on parity in the evolution of the language-ready brain, Aboitiz et al. (Aboitiz, 1995; Aboitiz et al., 2006; Aboitiz and Garcia, 1997) lay primary stress on the evolution of working memory systems. I see such alternatives as complementary, rather than either excluding the other.

With this, let me turn to a fuller exposition of the "Mirror System Hypothesis". I start with a few comparative comments concerning imitation to introduce key differences between monkey, ape and human that are relevant to understanding what such evolution may have involved. Monkeys have, at best, a very limited capacity for imitation (Visalberghi and Fragaszy, 1990; Voelkl and Huber, 2007), far overshadowed by what I call *simple imitation* as exhibited by apes. Myowa-Yamakoshi and Matsuzawa (1999) observed that chimpanzees took 12 or so trials to learn to "imitate" a behavior in a laboratory setting, focusing on bringing an object into relationship with another object or the body, rather than the actual movements involved. Byrne and Byrne (1993) found that gorillas learn complex feeding strategies but may take months to do so. Consider eating nettle leaves. Skilled gorillas grasp the stem firmly, strip off leaves, remove petioles bimanually, fold leaves over the thumb, pop the bundle into the mouth, and eat. The challenge for acquiring such skills is compounded because ape mothers seldom if ever correct and instruct their young (Tomasello, 1999) and because the sequence of "atomic actions" varies greatly from trial to trial. Byrne (2003) implicates *imitation by behavior parsing*, a protracted form of statistical learning whereby certain *subgoals* (e.g., nettles folded over the thumb) become evident from repeated observation as being common to most performances. In his account, the young ape may acquire the skill over many months by coming to recognize the relevant subgoals and derive action strategies for achieving them by trial-and-error.

This ability to learn the overall structure of a specific feeding behavior over many, many observations is very different from the human ability to understand any sentence of an open-ended set as it is heard, and generate another novel sentence as an appropriate reply. In many cases of praxis (i.e., skilled interaction with objects), humans need just a few trials to make sense of a relatively complex behavior if the constituent actions are familiar and the subgoals these actions must achieve are readily discernible, and they can use this perception to repeat the behavior under changing circumstances. We call this ability *complex imitation* (extending the definition of (Arbib, 2002) to incorporate the goal-directed imitation of Wohlschläger et al. (2003)). With such considerations in mind, I have elaborated the "Mirror System Hypothesis" (see (Arbib, 2005a) for a review, and commentaries on current controversies), defining an evolutionary progression of seven stages, S1 through S7:

- **S1:** Cortical control of hand movements.
- **S2:** A mirror system for grasping, shared with the common ancestor of human and monkey.

I stress that a mirror system does not provide imitation in itself. A monkey with an action in its repertoire may have mirror neurons active both when executing and observing that action yet does not repeat the observed action. Nor, crucially, does it use observation of a novel action to add that action to its repertoire. Thus, we hypothesize that evolution embeds a monkey-like mirror system in more powerful systems in the next two stages.

- **S3:** A simple imitation system for grasping, shared with the common ancestor of human and apes.
- **S4:** A complex imitation system for grasping which developed in the hominim line since that ancestor.

Each of these changes can be of evolutionary advantage in supporting the transfer of novel skills between the members of a community, involving praxis rather than explicit communication. We now explore the stages whereby our distant ancestors made the transition to *protolanguage*, in the sense of a communication system that supports the ready addition of new utterances by a group through some combination of innovation and social learning – it is open to the addition of new "protowords", in contrast to the closed set of calls of a group of nonhuman primates – yet lacks any tools, beyond mere juxtaposition of two or three protowords, to put protowords together to continually create novel utterances from occasion to occasion. Arbib et al. (submitted for publication), summarizing data on primate communication, note that monkey vocalizations are innately specified (though occasions for using a call may change with experience), whereas a group of apes may communicate with novel gestures, perhaps acquired by *ontogenetic ritualization* (Tomasello et al., 1997) whereby increasingly abbreviated and conventionalized form of an action may come to stand in for that action, an example being a beckoning gesture recognized by the child as standing for the parent's action of reaching out to grasp the child and pull it closer. This supports the hypothesis that it was gesture rather than vocalization (Seyfarth et al., 2005) that created the opening for greatly expanded gestural communication once complex imitation had evolved for practical manual skills. The expanded version of the "Mirror System Hypothesis" addresses this by positing the next two stages to be:

- **S5:** *Protosign*, a manual-based communication system breaking through the fixed repertoire of primate vocalizations to yield an open repertoire.
- **S6:** *Protolanguage as Protosign and Protospeech*: an expanding spiral of conventionalized manual, facial and vocal communicative gestures.

The transition from complex imitation and the small repertoires of ape gestures (perhaps 10 or so novel gestures shared by a group) to protosign involves, in more detail, first pantomime of grasping and manual praxic actions then of non-manual actions (e.g., flapping the arms to mime the wings of a flying bird), complemented by conventional gestures that simplify, disambiguate (e.g., to distinguish "bird" from "flying") or extend pantomime.

Pantomime transcends the slow accretion of manual gestures by ontogenetic ritualization, providing an "open semantics" for a large set of novel meanings (Stokoe, 2001). However, such pantomime is inefficient – both in the time taken to produce it, and in