

Although the view of Falk puts language emergence in a continuum that is closer to primate vocal communication than the framework of Michael Arbib, both models involve a progression atop the systems already preexisting in nonhuman primates. Arbib's work gives the first detailed account of putative evolutionary stages in the emergence of human language from a cognitive viewpoint. It therefore could be used as a framework to test specific links between cognitive human language and communicative human language emergence hypotheses, such as the one recently proposed by Falk.

Auditory object processing and primate biological evolution

Barry Horwitz,^a Fatima T. Husain,^a and Frank H. Guenther^b

^aBrain Imaging and Modeling Section, National Institute on Deafness and Other Communications Disorders, National Institutes of Health, Bethesda, MD 20892; ^bDepartment of Cognitive and Neural Systems, Boston University, Boston, MA 02215. horwitz@helix.nih.gov husainf@nidcd.nih.gov <http://www.nidcd.nih.gov/research/scientists/horwitzb.asp> guenther@cns.bu.edu <http://www.cns.bu.edu/~guenther/>

Abstract: This commentary focuses on the importance of auditory object processing for producing and comprehending human language, the relative lack of development of this capability in nonhuman primates, and the consequent need for hominid neurobiological evolution to enhance this capability in making the transition from protosign to protospeech to language.

The target article by Arbib provides a cogent but highly speculative proposal concerning the crucial steps in recent primate evolution that led to the development of human language. Generally, much of what Arbib proposes concerning the transition from the mirror neuron system to protosign seems plausible, and he makes numerous points that are important when thinking about language evolution. We especially applaud his use of neural modeling to implement specific hypotheses about the neural mechanisms mediating the mirror neuron system. We also think his discussion in section 6 of the necessity to use protosign as scaffolding upon which to ground symbolic auditory gestures in protospeech is a significant insight. However, the relatively brief attention Arbib devotes to the perception side of language, and specifically to the auditory aspects of this perception, seems to us to be a critical oversight. The explicit assumption that protosign developed before protospeech, reinforced by the existence of sign language as a fully developed language, allows Arbib (and others) to ignore some of the crucial features that both the productive and receptive aspects of speech require in terms of a newly evolved neurobiological architecture.

One aspect of auditory processing that merits attention, but is not examined by Arbib, has to do with auditory object processing. By auditory object, we mean a delimited acoustic pattern that is subject to figure-ground separation (Kubovy & Van Valkenburg 2001). Humans are interested in a huge number of such objects (in the form of words, melodic fragments, important environmental sounds), perhaps numbering on the order of 10^5 in an individual. However, it is difficult to train monkeys on auditory object tasks, and the number of auditory objects that interest them, compared to visual objects, seems small, numbering perhaps in the hundreds (e.g., some species-specific calls, some important environmental sounds). For example, Mishkin and collaborators (Fritz et al. 1999; Saunders et al. 1998) have showed that monkeys with lesions in the medial temporal lobe (i.e., entorhinal and perirhinal cortex) are impaired relative to unlesioned monkeys in their ability to perform correctly a visual delayed match-to-sample task when the delay period is long, whereas both lesioned and unlesioned monkeys are equally unable to perform such a task using auditory stimuli.

These results implicate differences in monkeys between vision and audition in the use of long-term memory for objects. Our view

is that a significant change occurred in biological evolution allowing hominids to develop the ability to discriminate auditory objects, to categorize them, to retain them in long-term memory, to manipulate them in working memory, and to relate them to articulatory gestures. It is only the last of these features that Arbib discusses. In our view, the neural basis of auditory object processing will prove to be central to understanding human language evolution. We have begun a systematic approach combining neural modeling with neurophysiological and functional brain imaging data to explore the neural substrates for this type of processing (Husain et al. 2004).

Concerning language production, Arbib's model of the mirror-neuron system (MNS) may require considerable modification, especially when the focus shifts to the auditory modality. For instance, there is no treatment of babbling, which occurs in the development of both spoken and sign languages (Petitto & Marientette 1991). Underscoring the importance of auditory processing in human evolution, hearing-impaired infants exhibit vocal babbling that declines with time (Stoel-Gammon & Otomo 1986).

However, there has been work in developing biologically plausible models of speech acquisition and production. In one such model (Guenther 1995), a role for the MNS in learning motor commands for producing speech sounds has been posited. Prior to developing the ability to generate speech sounds, an infant must learn what sounds to produce by processing sound examples from the native language. That is, he or she must learn an auditory target for each native language sound. This occurs in the model via a MNS involving speech sound-map cells hypothesized to correspond to mirror neurons (Guenther & Ghosh 2003). Only after learning this auditory target can the model learn the appropriate motor commands for producing the sound via a combination of feedback and feed-forward control subsystems. After the commands are learned, the same speech sound-map cell can be activated to read out the motor commands for producing the sound. In this way, mirror neurons in the model play an important role in both the acquisition of speaking skills and in subsequent speech production in the tuned system. This role of mirror neurons in development of new motor skills differs from Arbib's MNS model, which "makes the crucial assumption that the grasps that the mirror system comes to recognize are already in the (monkey or human) infant's repertoire" (sect. 3.2, para. 7).

Our efforts to comprehend the biological basis of language evolution will, by necessity, depend on understanding the neural substrates for human language processing, which in turn will rely heavily on comparative analyses with nonhuman primate neurobiology. All these points are found in Arbib's target article. A crucial aspect, which Arbib invokes, is the necessary reliance on neurobiologically realistic neural modeling to generate actual implementations of neurally based hypotheses that can be tested by comparing simulated data to human and nonhuman primate experimental data (Horwitz 2005). It seems to us that the fact that humans use audition as the primary medium for language expression means that auditory neurobiology is a crucial component that must be incorporated into hypotheses about how we must go beyond the mirror-neuron system.

On the neural grounding for metaphor and projection

Bipin Indurkha

International Institute of Information Technology, Hyderabad 500 019, India.
bipin@iiit.net

Abstract: Focusing on the mirror system and imitation, I examine the role of metaphor and projection in evolutionary neurolinguistics. I suggest that the key to language evolution in hominid might be an ability to project one's thoughts and feelings onto another agent or object, to see and feel things from another perspective, and to be able to empathize with another agent.

With regard to the evolutionary framework for neurolinguistics spelled out in Arbib's article, I would like to focus on the role of metaphor and projection therein. In particular, I am interested in the implications of Arbib's framework for the thesis "all knowledge (or language) is metaphorical." It should be clarified from the outset that this thesis is sometimes misconstrued to suggest that the literal or conventional does not exist – a suggestion that is trivially refuted. However, the sense in which I take it here is based on a well-known phenomenon that a novel metaphor sometimes becomes conventional through repeated use, and may even turn into polysemy; and the claim is that all that is conventional and literal now must have been metaphorical once (Indurkha 1994). Furthermore, I take the viewpoint that the key mechanism underlying metaphor, especially creative metaphor, is that of projection, which carves out a new ontology for the target of the metaphor (Indurkha 1992; 1998). This mechanism can be best explained as projecting a structure onto a stimulus, as in gestalt interaction, and is to be contrasted with the mapping-based approaches to metaphor, which require a pre-existing ontology for mapping. For example, in the context of Arbib's article, it is the projection mechanism that determines what constitutes objects and actions when a monkey watches a raisin being grasped by another monkey or by a pair of pliers.

There are two particular places in the evolutionary account articulated by Arbib where a projection step is implicit, and I shall zoom in on them in turn to raise some open issues. The first of these concerns the mirror neurons (sect. 3.2). Now, certain mirror neurons are known to fire when a monkey observes another monkey performing a particular grasping action but not when the grasp is being performed with a tool. This suggests a predisposition towards the ontology of a biological effector. The interesting question here is: How much variation can be introduced in the effector so that it is still acceptable to the mirror neuron. Does a robot arm trigger the mirror neuron? What about a hairy robot arm?

Similar remarks can be made with respect to the learning effect in mirror neurons. When a monkey first sees a raisin being grasped with a pair of pliers, then his mirror neurons do not fire. However, after many such experiences, the monkey's mirror neurons encoding precision grip start firing when he sees a raisin being grasped with pliers. This shows a predisposition towards the ontology of the object raisin and the effect of grip on it, as it is not the physical appearance of the effector but its effect on the object that matters. Again we may ask how much variation is possible in the object and the kind of grip before the mirror system fails to learn. For example, after the mirror neurons learn to fire on seeing a raisin being grasped with pliers, do they also fire when tweezers are used? Or, does the tweezers grasp have to be learned all over again?

These issues become more prominent when we consider imitation (sect. 4). In the literature, a wide range of animal behaviors are classified as imitation (Caldwell & Whiten 2002; Zentall & Akins 2001), and true imitation is distinguished from imprinting, stimulus enhancement, emulation learning, and so on. However, even in imitating a single action, one has to decide what aspect of the situation to imitate, as any situation has many possible aspects; and how to imitate, as the imitating agent has to interpret the situation from its point of view – it may not have the same effectors, access to the same objects, and so on – and project the observed action into its own action repertoire (Alissandrakis et al. 2002; Hofstadter 1995). In this respect, studies on the behavior of animals that imitate a non-conspecific model, such as bottlenose dolphins or parrots imitating a human model (or a bottlenose dolphin imitating a parrot?) are most illuminating. (See, e.g., Bauer & Johnson 1994; Kuczaj et al. 1998; Moore 1992.) In Arbib's framework, a distinction is made between simple and complex imitation to explain where humans diverge from monkeys, and a projection-like mechanism is posited for complex imitation (sect. 2.1: LR1; also sect. 5). But I would like to suggest that even simple imitation could invoke projection, and the crux of the distinction between humans and other animals might lie in the ability to interpret a

wider variety of actions and situations, and to project oneself into those situations to imitate them in a number of ways.

Empathy – being able to put oneself into another's shoes and to project one's thoughts and feelings into another person, animal, or object – is often considered a hallmark of being human. Indeed, one of the ideals of robotics research is to emulate this essentially human trait in robots. (See, e.g., Breazeal et al. 2005; Kozima et al. 2003. This is also the theme of the classic Philip K. Dick story "Do Androids Dream of Electric Sheep?" upon which the popular film *Blade Runner* was based.) A glimpse of the key role played by empathy in human cognition is provided by a study by Holstein (1970), in which children were given projection tasks such as being asked to imagine being a doorknob or a rock, and to describe one's thoughts and feelings in order to stimulate their creativity. In a very recent study, it was found that when participants hid one of their hands and a rubber hand was placed in front of them to make it look like their own hand, it took them only 11 seconds to project their feelings onto the rubber hand as if it were their own, down to the neural level: when the rubber hand was stroked by a brush, the somatosensory area in the participants' brain corresponding to their hand was stimulated (Ehrsson et al. 2004). One wonders if monkeys and other animals are capable of projecting their selves into other animals or other objects to this degree, and if the divergent point of hominid evolution might not be found therein.

Listen to my actions!

Jonas T. Kaplan and Marco Iacoboni

UCLA Brain Mapping Center, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095. jonask@ucla.edu
iacoboni@loni.ucla.edu <http://www.jonaskaplan.com>

Abstract: We believe that an account of the role of mirror neurons in language evolution should involve a greater emphasis on the auditory properties of these neurons. Mirror neurons in premotor cortex which respond to the visual and auditory consequences of actions allow for a modality-independent and agent-independent coding of actions, which may have been important for the emergence of language.

We agree with Arbib that the mirror property of some motor neurons most probably played an important role in the evolution of language. These neurons allow us to bridge the gap between two minds, between perception and action. As strong evidence for the role of mirror-like mechanisms in language, we have recently demonstrated with functional magnetic resonance imaging (fMRI) that a human cortical area encompassing primary motor and premotor cortex involved in the production of phonemes is also active during the perception of those same phonemes (Wilson et al. 2004). This suggests that motor areas are recruited in speech perception in a process of auditory-to-articulatory transformation that accesses a phonetic code with motor properties (Lieberman et al. 1967).

However, we direct our commentary mostly at what Arbib calls the transition from protosign to protospeech. In Arbib's account, a system of iconic manual gestures evolved from a mirror system of action recognition, and then somehow transitioned to a vocal-based language. Mention is made of the so-called audiovisual mirror neurons, which respond to the sound of an action as well as during the production of that action (Kohler et al. 2002). The role of these neurons in the evolution of language deserves more attention.

Arbib argues that arbitrary signs first evolved in gesture, which was more amenable to iconic representation, and that this protosign provided the "scaffolding" for vocal-based abstractions. We suggest that rather than being added on later, the auditory responsiveness of premotor neurons may have played a more central role in the development of abstract representations. The au-