that would allow us to discriminate between this and Arbib's hypotheses; the key desideratum is a better understanding of the neural basis of human vocal imitation (now sorely lacking).

The second stage I find problematic in Arbib's model is his explanation of the move from holistic protolinguistic utterances to analytic (fully linguistic) sentences. I agree that analytic models (which start with undecomposable wholes) are more plausible than synthetic models (e.g., Bickerton 2003; Jackendoff 1999) from a comparative viewpoint, because known complex animal signals map signal to meaning holistically. Both analytic and synthetic theories must be taken seriously, and their relative merits carefully examined. However, the robust early development of the ontogenetic "analytic insight" in modern human children renders implausible the suggestion that its basis is purely cultural, on a par with chess or calculus.

No other animal (including especially language-trained chimpanzees or parrots) appears able to make this analytic leap, which is a crucial step to syntactic, lexicalized language. While dogs, birds, and apes can learn to map between meanings and words presented in isolation or in stereotyped sentence frames, the ability to extract words from arbitrary, complex contexts and to recombine them in equally complex, novel contexts is unattested in any nonhuman animal. In vivid contrast, each generation of human children makes this "analytic leap" by the age of three, without tutelage, feedback, or specific scaffolding. This is in striking contrast to children's acquisition of other cultural innovations such as alphabetic writing, which occurred just once in human history and still poses significant problems for many children, even with long and detailed tutelage.

Although the first behavioural stages in the transition from holistic to analytic communication were probably Baldwinian exaptations, they must have been strongly and consistently shaped by selection since that time, given the communicative and conceptual advantages that a compositional, lexicalized language offers. The "geniuses" making this analytic insight were not adults, but children, learning and (over)generalizing about language unanalyzed by their adult caretakers, and this behaviour must have been powerfully selected, and genetically canalized, in recent human evolution. It therefore seems strange and implausible to claim that the acquisition of the analytic ability had "little if any impact on the human genome" (target article, sect. 2.3).

In conclusion, by offering an explicit phylogenetic hypothesis, detailing each hypothetical protolinguistic stage and its mechanistic underpinnings, and allowing few assumptions about these stages to go unexamined, Arbib does a service to the field, goes beyond previous models, and raises the bar for all future theories of language phylogeny. However, further progress in our understanding of language evolution demands parallel consideration of multiple plausible hypotheses, and finding empirical data to test between them, on the model of physics or other natural sciences. Arbib's article is an important step in this direction.

Imitation systems, monkey vocalization, and the human language

Emmanuel Gilissen

Royal Belgian Institute of Natural Sciences, Anthropology and Prehistory, B-1000 Brussels, Belgium. **Emmanuel.Gilissen@naturalsciences.be**
http://www.naturalsciences.be

Abstract: In offering a detailed view of putative steps towards the emergence of language from a cognitive standpoint, Michael Arbib is also introducing an evolutionary framework that can be used as a useful tool to confront other viewpoints on language evolution, including hypotheses that emphasize possible alternatives to suggestions that language could not have emerged from an earlier primate vocal communication system.

An essential aspect of the evolutionary framework presented by Michael Arbib is that the system of language-related cortical areas evolved atop a system that already existed in nonhuman primates. As explained in the target article, crucial early stages of the progression towards a language-ready brain are the mirror system for grasping and its extension to permit imitation.

When comparing vocal-acoustic systems in vertebrates, neuroanatomical and neurophysiological studies reveal that such systems extend from forebrain to hindbrain levels and that many of their organizational features are shared by distantly related vertebrate taxa such as teleost fish, birds, and mammals (Bass & Baker 1997; Bass & McKibben 2003; Goodson & Bass 2002). Given this fundamental homogeneity, how are documented evolutionary stages comparable to imitation in vertebrate taxa? Vocal imitation is a type of higher-level vocal behaviour that is, for instance, illustrated by the songs of humpback whales (Payne & Payne 1985). In this case, there is not only voluntary control over the imitation process of a supposedly innate vocal pattern, but also a voluntary control over the acoustic structure of the pattern.

This behaviour seems to go beyond "simple" imitation of "object-oriented" sequences and resembles a more complex imitation system. Although common in birds, this level of vocal behaviour is only rarely found in mammals (Jürgens 2002). It "evolved atop" preexisting systems, therefore paralleling emergence of language in humans. It indeed seems that this vocalization-based communication system is breaking through a fixed repertoire of vocalizations to yield an open repertoire, something comparable to protosign stage (S5). Following Arbib, S5 is the second of the three stages that distinguish the hominid lineage from that of the great apes. Although the specific aspect of S5 is to involve a manual-based communication system, it is interesting to see how cetaceans offer striking examples of convergence with the hominid lineage in higher-level complex cognitive characteristics (Marino 2002).

The emergence of a manual-based communication system that broke through a fixed repertoire of primate vocalizations seems to owe little to nonhuman primate vocalizations. Speech is indeed a learned motor pattern, and even if vocal communication systems such as the ones of New World monkeys represent some of the most sophisticated vocal systems found in nonhuman primates (Snowdon 1989), monkey calls cannot be used as models for speech production because they are genetically determined in their acoustic structure. As a consequence, a number of brain structures crucial for the production of learned motor patterns such as speech production are dispensable for the production of monkey calls (Jürgens 1998).

There is, however, one aspect of human vocal behavior that does resemble monkey calls in that it also bears a strong genetic component. This aspect involves emotional intonations that are superimposed on the verbal component. Monkey calls can therefore be considered as an interesting model for investigating the central mechanisms underlying emotional vocal expression (Jürgens 1998).

In recent studies, Falk (2004a; 2004b) hypothesizes that as human infants develop, a special form of infant-directed speech known as baby talk or motherese universally provides a scaffold for their eventual acquisition of language. Human babies cry in order to re-establish physical contact with caregivers, and human mothers engage in motherese that functions to soothe, calm, and reassure infants. These special vocalizations are in marked contrast to the relatively silent mother/infant interactions that characterize living chimpanzees (and presumably their ancestors). Motherese is therefore hypothesized to have evolved in early hominin mother/infant pairs, and to have formed an important prelinguistic substrate from which protolanguage eventually emerged. Although we cannot demonstrate whether there is a link between monkey calls and motherese, it appears that the neural substrate for emotional coding, prosody, and intonation, and hence for essential aspects of motherese content, is largely present in nonhuman primate phonation circuitry (Ploog 1988; Sutton & Jürgens 1988). In a related view, Deacon (1989) suggested that the vocalization circuits that play a central role in nonhuman primate vocalization became integrated into the more distributed human language circuits.

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]

[a]*Brain Imaging and Modeling Section, National Institute on Deafness and Other Communications Disorders, National Institutes of Health, Bethesda, MD 20892;* [b]*Department 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 & Marentette 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 Indurkhya

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.