cesses, therefore the complexity of the imitation system of an organism cannot exceed the complexity of the systems to be imitated. This principle seriously constrains the possibility of the emergence of a new, more complex imitation system without the corresponding complicating within the systems to be imitated. Such a possibility seems to underlie Arbib's approach because, in emphasizing the changes in the imitation system, he does not require similar fundamental changes in other systems.

Of course, it is impossible to abandon the idea that the complex imitation system could emerge as a result of a single mutation without the corresponding changes in other systems of some ancient hominids; but such hominids occasionally benefited from their new possibilities, thereby surviving successfully, until other systems achieved the complexity of the imitation system; and then natural selection started working more conventionally again. The probability of this scenario is extremely low, obviously. Another approach to the origin of the complex imitation system, which seems much more probable, is that a certain complication of other systems preceded this system and made its appearance necessary. This, however, means that Arbib's hypothesis suggesting that the complex imitation system is the "missing link" is not correct, because other systems in fact determined the appearance of language.

Like other hypotheses of language origin, Arbib's hypothesis is based on the idea that language is a means of communication. This definition is correct but incomplete: language is a means of communication for people engaged in a joint activity. There is a clear correlation between the diversity of activities and the complexity of the language serving these activities. Modern languages consist of hundreds of thousands of words only because these languages are applied in thousands of diverse activities. Each human activity is goal-directed, hence, the complexity of languages is a consequence of the ability of the human brain to construct diverse goals. Indeed, most human goals are not constrained by any innate basis; they are social, and result from interactions between people. So, there is an obvious connection between language and the ability to construct and maintain long-term motivations with no innate basis.

No nonhuman animals have a motivational system with similar characteristics. Animals have long-term motivations (e.g., sex, hunger), but these are all innate. An animal can form learned motivations, but only when its basic drives are activated. The hypothesis that the motivation of animals is always constrained by the activation of basic drives was suggested by Kohler (1917/1927), and despite intensive researches, there have still been no data inconsistent with it (Suddendorf & Corballis 1997). With the limited and stable number of long-term motivations, animals are constrained in using and developing their languages. Since all their motivations are connected with vital functions, any serious misunderstanding in the process of communication can be fatal; as a result, the number of signals in animal languages must be limited, and the signals must have unequivocal meanings. Roughly speaking, animals do not have a language similar to human languages because they simply do not need it.

I have suggested elsewhere that the emergence of the ability to construct and maintain long-term goals with no innate basis was the missing link for language (Prudkov 1999c) and for the other distinctively human characteristics (Prudkov 1999a; 1999b) because the ability allowed ancient humans to overcome the constraints of innate motivations, thus providing the possibility of constructing new, flexible, and open systems. In other words, protolanguage emerged because in new situations conditioned by goals having no innate basis, the innate communicative means became inefficient for interactions between ancient hominids, and those who were able to construct new means succeeded in reproduction. Of course, language, imitation, and the theory of mind had started evolving then. It is very important to emphasize that without the prior (or parallel) formation of the system able to construct learned, long-term motivations, any changes in other systems (e.g., in intelligence) were not sufficient to overcome innate constraints. For example, the capacity of birds to navigate in three-dimensional space on the basis of visual cues obviously exceeds that of humans, but innate mechanisms determine the behavior of birds.

It is reasonable to think that there was a reciprocal interaction in the evolution of human language and motivation. The new motivational ability spurred the development of language; afterwards language was used to construct efficient, purposeful processes, and this interaction likely determined all stages of human evolution. This joint evolution was facilitated by the fact that a common mechanism that evolved within these systems is the capacity to form and execute complex, hierarchical, goal-directed processes (such processes are rapid and relatively simple in language and are slow and complex in motivation) (Prudkov & Rodina 1999). In other words, I agree with Arbib that humans have a language-ready brain rather than special mechanisms embedded in the genome. The capacity was also involved in the development of the imitation system, because a basic characteristic distinguishing the human imitation system from its animal analogs is the possibility to imitate more complex and long-term processes. But the development of the imitation system itself is not sufficient to construct protolanguage, because only the new motivational system could make imitation voluntary and arbitrary. Indeed, in emphasizing that at a certain stage of evolution communication became voluntary and intentional, Arbib does not explain what mechanisms underlay such possibilities of communication.

In my opinion, the gestural and vocal components of protolanguage emerged together, but the latter gained advantage in the development because, unlike gestures, which are effective only in dyadic contacts, vocalizations are more effective in group actions (group hunting, collective self-defense, etc.), which became the first actions guided by goals having no innate basis.

Vocal gestures and auditory objects

Josef P. Rauschecker

Laboratory of Integrative Neuroscience and Cognition, Georgetown University School of Medicine, Washington, DC 20057-1460.
rauschej@georgetown.edu

Abstract: Recent studies in human and nonhuman primates demonstrate that auditory objects, including speech sounds, are identified in anterior superior temporal cortex projecting directly to inferior frontal regions and not along a posterior pathway, as classically assumed. By contrast, the role of posterior temporal regions in speech and language remains largely unexplained, although a concept of vocal gestures may be helpful.

In his target article, Arbib maintains (and before him, Rizzolatti & Arbib 1998) that language originated from a system of mirror neurons coding manual gestures, rather than from vocal communication systems present in nonhuman primates (and other animals). I do not doubt the usefulness of the mirror-neuron concept, which brings back to mind the motor theory of speech perception (Liberman et al. 1967). In fact, many recent neuroimaging studies have independently demonstrated a simultaneous activation of what were previously thought of as separate centers for the production and perception of human language, Broca's and Wernicke's areas, respectively. These designations go back more than a century to crudely characterized single-case studies of neurological patients, which have been shown by modern magnetic resonance imaging (MRI) techniques (Bookheimer 2002) to have missed much more brain than the relatively small regions that now bear their discoverers' names.

Both on that basis and on the basis of his own belief in intertwined systems of perception and action, it is surprising that Arbib continues to use this outdated terminology. "Broca's area" at least is redefined by him as part of a system that deals with, among others, "sequential operations that may underlie the ability to

https://doi.org/10.1017/S0140525X05410031 Published online by Cambridge University Press

form words out of dissociable elements" (sect. 8), a definition that many researchers could agree with, although the exact correspondence with cytoarchitectonically defined areas and the homologies between human and nonhuman primates are still controversial. "Wernicke's area," by contrast, gets short shrift. Arbib talks about it as consisting of the posterior part of Brodmann's area 22, including area Tpt of Galaburda and Sanides (1980) and an "extended [parietal area] PF," suggesting that this is the only route that auditory input takes after it reaches primary auditory cortex. Of course, this suggestion echoes the classical textbook view of a posterior language pathway leading from Wernicke's to Broca's area via the arcuate fascicle.

A remarkable convergence of recent neurophysiological and functional imaging work has demonstrated, however, that the analysis of complex auditory patterns and their eventual identification as auditory objects occurs in a completely different part of the superior temporal cortex, namely, its anterior portion. The anterior superior temporal (aST) region, including the anterior superior temporal gyrus (STG) and to some extent the dorsal aspect of the superior temporal sulcus (STS), project to the inferior frontal (IF) region and other parts of the ventrolateral prefrontal cortex (VLPFC) via the uncinate fascicle. Together, the aST and IF cortices seem to form a "what" stream for the recognition of auditory objects (Rauschecker 1998; Rauschecker & Tian 2000), quite similar to the ventral stream for visual object identification postulated previously (Ungerleider & Mishkin 1982). Neurophysiological data from rhesus monkeys suggest that neurons in the aST are more selective for species-specific vocalizations than are neurons in the posterior STG (Tian et al. 2001). In humans, there is direct evidence from functional imaging work that intelligible speech as well as other complex sound objects are decoded in the aST (Binder et al. 2004; Scott et al. 2000; Zatorre et al. 2004).

It seems, therefore, that the same anatomical substrate supports both the decoding of vocalizations in nonhuman primates and the decoding of human speech. If this is the case, the conclusion is hard to escape that the aST in nonhuman primates is a precursor of the same region in humans and (what Arbib may be reluctant to accept) that nonhuman primate vocalizations are an evolutionary precursor to human speech sounds. Indeed, the same phonological building blocks (or "features"), such as frequency-modulated (FM) sweeps, band-passed noise bursts, and so on, are contained in monkey calls as well as human speech. Admittedly, the decoding of complex acoustic sound structure alone is far from sufficient for language comprehension, but it is a necessary precondition for the effective use of spoken speech as a medium of communication. Arbib argues, with some justification, that communication is not bound to an acoustic (spoken) medium and can also function on the basis of visual gestures. However, in most hearing humans the acoustic medium, that is, "vocal gestures," have gained greatest importance as effective and reliable carriers of information.

An interesting question remaining, in my mind, is, therefore, how the auditory feature or object system in the aST could interact with a possible mirror system, as postulated by Arbib and colleagues. The projection from aST to IF seems like a possible candidate to enable such an interaction. Indeed, auditory neurons, some of them selectively responsive to species-specific vocalizations, are found in the VLPFC (Romanski & Goldman-Rakic 2002). According to our view, aST serves a similar role in the auditory system as inferotemporal (IT) cortex does for the visual system. Which role, if any, Wernicke's area (or posterior STG) plays for vocal communication, including speech and language, remains the bigger puzzle. Understanding it as an input stage to parietal cortex in an auditory dorsal pathway is a good hint. However, as Arbib would say, "empirical data are sadly lacking" and need to be collected urgently.

Continuities in vocal communication argue against a gestural origin of language

Robert M. Seyfarth

Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104. **seyfarth@psych.upenn.edu**
http://www.psych.upenn.edu/~seyfarth/Baboon%20research/index.htm

Abstract: To conclude that language evolved from vocalizations, through gestures, then back to vocalizations again, one must first reject the simpler hypothesis that language evolved from prelinguistic vocalizations. There is no reason to do so. Many studies – not cited by Arbib – document continuities in behavior, perception, cognition, and neurophysiology between human speech and primate vocal communication.

Arbib argues that the emergence of human speech "owes little to nonhuman vocalizations" and concludes that "evolution did not proceed directly from monkey-like primate vocalizations to speech but rather proceeded from vocalization to manual gesture and back to vocalization again" (target article, sect. 2.3). Accepting this hypothesis requires us to adopt a convoluted argument over a simple one. There is no need to do so.

If dozens of scientists had been studying the natural vocalizations of nonhuman primates for the past 25 years and all had concluded that the vocal communication of monkeys and apes exhibited no parallels whatsoever with spoken language, one might be forced to entertain Arbib's hypothesis. If years of neurobiological research on the mechanisms that underlie the perception of calls by nonhuman primates had revealed no parallels with human speech perception, this, too, might compel us to reject the idea that human language evolved from nonhuman primate vocalizations. Neither of these conclusions, however, is correct.

Arbib offers his hypothesis as if he had carefully reviewed the literature on nonhuman primate vocal communication and thoughtfully rejected its relevance to the evolution of human language. Readers should be warned, however, that his review ends around 1980 and even neglects some important papers published before that date.

Primate vocal repertoires contain several different call types that grade acoustically into one another. Despite this inter-gradation, primates produce and perceive their calls as, roughly speaking, discretely different signals. Different call types are given in different social contexts (e.g., Cheney & Seyfarth 1982; Fischer 1998; Fischer et al. 2001a; Hauser 1998; Snowdon et al. 1986). In playback experiments, listeners respond in distinct ways to these different call types, as if each type conveys different information (e.g., Fischer 1998; Fischer et al. 2001b; Rendall et al. 1999). Listeners discriminate between similar call types in a manner that parallels – but does not exactly duplicate – the categorical perception found in human speech (Fischer & Hammerschmidt 2001; Owren et al. 1992; Prell et al. 2002; Snowdon 1990; Zoloth et al. 1979). Offering further evidence for parallels with human speech, the grunts used by baboons (and probably many other primates) differ according to the placement of vowel-like formants (Owren et al. 1997; Rendall 2003).

Arbib incorrectly characterizes primate vocalizations as "involuntary" signals. To the contrary, ample evidence shows that nonhuman primate call production can be brought under operant control (Peirce 1985) and that individuals use calls selectively in the presence of others with whom they have different social relations (for further review and discussion, see Cheney & Seyfarth 1990; Seyfarth & Cheney 2003b).

Because nonhuman primates use predictably different calls in different social and ecological contexts, listeners can extract highly specific information from them, even in the absence of any supporting contextual cues. For example, listeners respond to acoustically different alarm calls as if they signal the presence of different predators (Fichtel & Hammerschmidt 2002; Fischer 1998; Seyfarth et al. 1980), and to acoustically different grunts as if they signal the occurrence of different social events (Cheney & Sey-

https://doi.org/10.1017/S0140525X05410031 Published online by Cambridge University Press