

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-

farth 1982; Rendall et al. 1999). In habituation-dishabituation experiments that asked listeners to make a same-different judgment between calls, subjects assessed calls based on their meaning, not just their acoustic properties (Cheney & Seyfarth 1988; Zuberbuhler et al. 1999). The parallels with children's perception of words cannot be ignored (see Zuberbuhler 2003 for review).

Indeed, it is now clear that although primates' production of vocalizations is highly constrained, their ability to extract complex information from sounds is not (Seyfarth & Cheney 2003b). Upon hearing a sequence of vocalizations, for example, listeners acquire information that is referential, discretely coded, hierarchically structured, rule-governed, and propositional (Bergman et al. 2003; Cheney & Seyfarth, in press). These properties of primates' social knowledge, although by no means fully human, bear striking resemblances to the meanings we express in language, which are built up by combining discrete-valued entities in a structured, hierarchical, rule-governed, and open-ended manner. Results suggest that the internal representations of language meaning in the human brain initially emerged from our prelinguistic ancestors' knowledge of social relations, as exhibited in the information they acquire from vocalizations (Cheney & Seyfarth 1997; in press; Worden 1998).

Nonhuman primate vocalizations also exhibit parallels with human speech in their underlying neural mechanisms. Behavioral studies of macaques suggest that the left hemisphere is specialized for processing species-specific vocalizations but not other auditory stimuli (Hauser & Anderson 1994; Petersen et al. 1978). Lesion results demonstrate that ablation of auditory cortex on the left but not the right hemisphere disrupts individuals' ability to discriminate among acoustically similar call types (Heffner & Heffner 1984). Most recently, Poremba et al. (2004) measured local cerebral metabolic activity as macaques listened to a variety of auditory stimuli. They found significantly greater activity in the left superior temporal gyrus as compared with the right, but only in response to conspecific vocalizations. These and other results (e.g., Wang et al. 1995; see Hauser [1996] and Ghazanfar & Hauser [2001] for review) suggest that Arbib is wrong to assume that primate vocalizations "appear to be related to non-cortical regions" (sect. 1. 2, para. 3). They further suggest that the neurophysiological mechanisms underlying human speech processing evolved from similar mechanisms in our nonhuman primate ancestors.

In sum, research demonstrates a striking number of continuities – in behavior, perception, cognition, and neurophysiology – between human speech and the vocal communication of nonhuman primates. Nonhuman primate vocal communication does not qualify as language, but it does exhibit many of the characteristics that one would expect to find if human language had evolved from the vocal communication and cognition of the common ancestor of human and nonhuman primates.

Arbib cites none of this research. As a result, his presentation is strongly biased in favor of his own view that the emergence of human speech "owes little to nonhuman vocalizations" (target article, Abstract). To accept the convoluted hypothesis that spoken language evolved from vocalizations, through gestures, then back to vocalizations again, one must first have good reason to reject the simpler hypothesis that spoken language evolved from prelinguistic vocal communication. A substantial body of data argues against such a rejection.