

Noble & Davidson 1996). Mind, consciousness, and, especially, capacity for and realization of culture constitute, at least in part, neuroanatomical and neurophysiological phenomena. As the hominid brain evolved, episodic and, especially, semantic memory contained material that was fed into a working memory bin or supervisory system providing basis for experience and (autobiographical?) selfhood (Baddeley 1986; Fuster 2002; Tononi & Edelman 1998; Shallice 1988).

When evolutionary scientists address such topics, they focus on concrete, expedient, raw, or “brutish” fitness imperatives, involving such things as hunting, foraging, mating, or parenting (Wray 1998; 2000), leaving out cultural, symbolic, ritual complexities (Fabrega 1997; 2002; 2004; Knight 1991). Arbib has managed to touch on all of these matters implicitly and tangentially, but for the most part leaves them off his MNS stage.

Beginning with the language-readiness phases wherein intended communication is explicitly manifest, particularly during the shift from imitation to (conscious use of) protosign, then to protospeech, and finally to language, Arbib insinuates (and once mentions) culture/community and implies a sense of shared social life and social history. *If* there is a shared body of knowledge about what pantomimes are for and what they mean, what disambiguating gestures are for and mean, and what speech sounds are for and mean, *then* there exists an obvious meaning-filled thought-world or context “carried in the mind” that encompasses self-awareness, other-awareness, need for cooperation, capacity for perspective-taking – and, presumably, a shared framework of what existence, subsistence, mating, parenting, helping, competing, and the like entail and what they mean. All of this implies that evolution of culture “happened” or originated during phases of biological evolution as LR capacities came into prominence (Foley 2001). No one denies that “culture” was evident at 40,000 B.C.E., yet virtually no one ventures to consider “culture” prior to this “explosion.” Arbib implies, along with Wray (1998; 2000) that the context of language evolution was dominated by purely practical, expedient considerations (e.g., getting things done, preserving social stability, greetings, requests, threats). Boyer (1994) and Atran and Norenzayan (2004) imply that as a human form of cognition “coalesced,” so did a significant component of culture (Fabrega 1997; 2002; 2004). Arbib’s formulation *suggests* culture “got started” well before this, perhaps, as he implies, with *Homo habilis* and certainly *Homo erectus*.

Protomusic and protolanguage as alternatives to protosign

W. Tecumseh Fitch

School of Psychology, University of St. Andrews, Fife KY16 9JP, Scotland.
wtsf@st-andrews.ac.uk

Abstract: Explaining the transition from a signed to a spoken protolanguage is a major problem for all gestural theories. I suggest that Arbib’s improved “beyond the mirror” hypothesis still leaves this core problem unsolved, and that Darwin’s model of musical protolanguage provides a more compelling solution. Second, although I support Arbib’s analytic theory of language origin, his claim that this transition is purely cultural seems unlikely, given its early, robust development in children.

Arbib’s wide-ranging paper commendably weaves together multiple threads from neuroscience, linguistics, and ethology, providing an explicit, plausible model for language phylogeny, starting with our common ancestor with other primates and ending with modern language-ready *Homo sapiens*. He takes seriously the comparative data accrued over the last 40 years of primatology, rightly rejecting any simple transition from “monkey calls to language,” and provides an excellent integrative overview of an important body of neuroscientific data on grasping and vision and their interaction. I agree with Arbib’s suggestion that some type of “protolanguage” is a necessary stage in language evolution, and

that the term should not be limited to any particular model of protolanguage (e.g., Bickerton’s [1995] model). However, I suggest that the relevance of monkey mirror neurons to gestural theories of language evolution has been overstated, and I will focus on weaknesses Arbib’s model faces in explaining two key transitions: protosign to protospeech, and holistic protolanguage to syntactic language.

The chain of a logical argument is only as strong as its weakest link. The weak link in Arbib’s model is the crucial leap from protosign to protospeech, specifically his elision between two distinct forms of imitation: vocal and manual. Comparative data suggest that these two are by no means inevitably linked. Although dolphins are accomplished at both whole-body and vocal imitation (Reiss & McCowan 1993; Richards et al. 1984), and parrots can imitate movements (Moore 1992), evidence for non-vocal imitation in the largest group of vocal imitators, the songbirds, is tenuous at best (Weydemeyer 1930). Apes exhibit the opposite dissociation between some manual proto-imitation with virtually no vocal imitation. There is therefore little reason to assume that the evolution of manual imitation and protosign would inevitably “scaffold” vocal imitation. Realizing this, Arbib offers a neuroanatomical justification for this crucial link, suggesting that the hypertrophied manual mirror system supporting protosign “colonized” the neighboring vocal areas of F5 by a process of “collateralization.”

However, the key difference between human and other primate brains is not limited to local circuitry in area F5 but includes long-distance corticomotor connections from (pre)motor cortex to auditory motor neurons in the brainstem, which exist in humans but not other primates (Jürgens 1998). These probably represent a crucial neural step in gaining the voluntary control over vocalization differentiating humans from monkeys and apes. “Collateralization” is not enough to create such corticomotor connections. Indeed, given competition for cortical real estate in the developing brain, it would seem, if anything, to make their survival *less* likely. Thus, like other versions of gestural origins hypotheses, Arbib’s model fails to adequately explain how a “protosign” system can truly scaffold the ability for vocal learning that spoken language rests upon. Are there alternatives?

Darwin suggested that our prelinguistic ancestors possessed an intermediate “protolanguage” that was more musical than linguistic (Darwin 1871). Combining Darwin’s idea with the “holistic protolanguage” arguments given by Arbib and others (Wray 2002a), and the “mimetic stage” hypothesized by (Donald 1993), gives a rather different perspective on the co-evolution of vocal and manual gesture, tied more closely to music and dance than pantomime and linguistic communication. By this hypothesis, the crucial first step in human evolution from our last common ancestor with chimpanzees was the development of vocal imitation, similar in form and function to that independently evolved in many other vertebrate lineages (including cetaceans, pinnipeds, and multiple avian lineages). This augmented the already-present movement display behaviour seen in modern chimpanzees and gorillas to form a novel, learned, and open-ended multimodal display system. This hypothetical musical protolanguage preceded any truly linguistic system capable of communicating particulate, propositional meanings.

This hypothesis is equally able to explain the existence of sign (via the dance/music linkage), makes equal use of the continuity between ape and human gesture, and can inherit all of Arbib’s “expanding spiral” arguments. But it replaces the weakest link in Arbib’s logical chain (the scaffolding of vocal by manual imitation) with a step that appears to evolve rather easily: the early evolution of a vocally imitating “singing ape” (where vocal learning functions in enhancement of multimodal displays). It renders understandable why all modern human cultures choose speech over sign as the linguistic medium, if this sensory-motor channel is available. It also explains, “for free,” the evolution of two nonlinguistic human universals, dance and music, as “living fossils” of an earlier stage of human communicative behaviour. We need posit no hypothetical or marginal protolanguage: evidence of a human-specific music/dance communication system is as abundant as one could desire. There are abundant testable empirical predictions

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 ar-

eas 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.