group uses it. However, the speed of this process critically depends on such institutions as writing, hierarchical social organisation (the most powerful accelerator of social development; Cavalli-Sforza & Feldman 1981), and at least rudimentary mass media. Churches and monasteries played an active role in dissemination of new notions and concepts in Europe as well as the Far East.

Arbib argues that the development of modern languages such as English required much less time than the time to pass over from protolanguage to language. This analogy misses, however, the simple fact that modern languages did not start with a protolanguage. Rather, their starting point was another highly developed language. Italian needed only 800 years to reach its peak in *The Divine Comedy*, but its precursor was Latin.

More generally, the problem can be formulated as follows: the proposed theory postulates that the development of language was not supported by natural selection. But the major social mechanisms (e.g., the mechanisms of state, church, writing, social hierarchies, and fast migration), which might be supposed to have replaced evolutionary mechanisms, did not exist when first languages developed from their protolanguage ancestors. On the other hand, social mechanisms which were present from the very beginning (e.g., socialization in tribes and family education) are known to be factors of conservation rather than development. Due to these social processes I would expect that genial inventors of words were ostracized rather than accepted. Hence, it remains unclear how, if we retain Arbib's example, the new notion "sour" might ever have become known to anybody except the closest fellows of its genial inventor. Therefore, any generalisation about the development of the first human language(s) from what is known about modern languages is problematic.

Given that the degrees of linguistic and genetic similarity between populations correlate (Cavalli-Sforza 1996), and that the transition from protolanguage to language can have covered 1,500 to 2,000 generations, I do not understand why biological mechanisms should be denied during the evolvement of the very first (but not proto-) language. A possible argument could be the lack of substantial biological progress between the early *Homo sapiens*, having only a protolanguage, and modern people. But this argument would be misleading because it confounds evolution with progress and power of different brains with their diversity. There was not a big genetic progress since the appearance of Homo sapiens, but the genetic changes took place.

ACKNOWLEDGMENT

This work was partially supported by a grant from the Fortune Foundation, University of Tübingen Medical School.

NOTE

1. From the pragmatic point of view, a message always remains "here and now." For instance, I am going to discuss the transition from protolanguage to language, which was about 100,000 years ago, that is, fairly "beyond the here-and-now"; but my aim is to convince Arbib or other readers today.

Evolutionary sleight of hand: Then, they saw it; now we don't

Peter F. MacNeilage[a] and Barbara L. Davis[b]

[a]*Department of Psychology, The University of Texas at Austin, A8000 Austin, TX 78712;* [b]*Department of Communication Sciences and Disorders, The University of Texas at Austin, A1100 Austin, TX 78712.*

macneilage@psy.utexas.edu **babs@mail.utexas.edu**

Abstract: Arbib's gestural-origins theory does not tell us why or how a subsequent switch to vocal language occurred, and shows no systematic concern with the signalling affordances or constraints of either medium. Our frame/content theory, in contrast, offers both a *vocal* origin in the invention of kinship terms in a baby-talk context and an explanation for the structure of the currently favored medium.

Why is there such a continued interest in formulating gestural-origins theories of language when they never provide an adequate reason for the subsequent abandonment of the gestural medium, or a means of getting us to the eventual vocal one? As to *why* the change occurred, Arbib finesses that issue. The usual explanations – that signed language is not omnidirectional, does not work in the dark, and ties up the hands – have always constituted an insufficient basis for such a radical reorganization. As to *how* the change occurred, we note that the first gestural-origins theory of the modern era was proposed by Hewes (1973; 1996), who gracefully admitted that "The ideas about the movement from a postulated prespeech language to a rudimentary spoken one are admittedly the weakest part of my model" (1996, p. 589). Nothing has changed since, whether in Arbib's earlier gestural incarnation (Arbib & Rizzolatti 1997), in the most recent reincarnation of Corballis's gestural-origins theory (Corballis 2003a; see MacNeilage 2003 for commentary), or in the present target article.

Arbib is more vulnerable than most on the *why* problem because he posits an original *open* (read unrestricted) pantomimic protosign stage. Openness is a definitional property of true language. Hockett (1978) pointed out, we think correctly, that if manual communication had ever achieved openness, this would have been such a momentous development that we would never have abandoned the original form of the incarnation. Besides ignoring the *why* question, Arbib palms the *how* question, saying only "Once an organism has an iconic gesture, it can both modulate that gesture and/or or symbolize it (non-iconically) by 'simply' associating a vocalization with it" (sect. 6.1, para. 2, Arbib's quotation marks). Simply?

Arbib's problems arise from a very disappointing source, given his own focus on the evolution of action. He shows little regard for the affordances and constraints of the two language transmission media (their action components). He consequently misses a number of opportunities to put constraints on his model. For example, his problematical conclusion that pantomime could be an open system disregards a commonly accepted conclusion in linguistics that for language to become an open system, it must have a combinatorial phonology consisting of meaningless elements (such as consonants and vowels in the vocal medium, and hand shapes, locations, and movements in the manual medium) (Jackendoff 2002; Studdert-Kennedy & Lane 1980). He makes scant reference to modern-day sign languages, apparently regarding them as an adventitious side effect rather than a central phenomenon that must be accounted for in a language-evolution context. Where did modern day sign languages get the combinatorial phonology commonly thought to be necessary for an open linguistic system, if their predecessor already had an open pantomimic system? Arbib says nothing about the system-level problems of getting from a pantomimic repertoire to a speech repertoire at either the perceptual or the motor level.

A prominent consequence of Arbib's neglect of the linguistic action component is shown in his dubious contention that hominids in the protospeech stage could have dashed off complex semantic concepts with holistic phonetic utterances such as "grooflack" or "koomzash," forms that take a modern infant several years to master. Such utterances are not holistic today. How could forms with such internal complexity, sounding like words with modern structure, have originated, and how could they have become linked with concepts? Also, if they indeed existed as holistic complexes, as Arbib claims, how did they get fractionated? And how was the phonetic fractionation related to the putative semantic fractionation into present-day forms of class elements such as nouns and verbs in a way that is consistent with phonology-morphology relationships in present-day languages?

In light of the problems of gestural origins theories with the *why* and *how* questions, there is a need for a theory of evolution of language that gets us to modern language in the old-fashioned way – by speaking it! Our frame/content theory (MacNeilage 1998; MacNeilage & Davis 1990; 2000) is such a theory. Arbib bills our theory as being about "the evolution of syllabification as a way

to structure vocal gestures" but asserts that it "offers no clue as to what might have linked such a process to the expression of meaning" (sect. 6.1, para. 3). Apparently, Arbib did not revise the target article following an exchange of critiques with him earlier this year (our paper not being cited in the target article), in which we described our view that the first words may have been kinship terms formed in the baby-talk context. (For this exchange, see Arbib 2005; MacNeilage & Davis, in press b.)

Our primary contribution in this regard has been to refine earlier conceptions (cf. Locke 1993) of exactly how kinship terms might have originated in a baby-talk context (MacNeilage & Davis 2004; in press a). Our argument is that the structure of present-day baby-talk words is basically identical to the structure of the first words of early speakers of language. We propose that because of this basic identity, the first words had forms like baby-talk forms.

The basic idea (see Falk 2004a, for a recent version) starts from the contention that nasal vocalizations of infants in the presence of the mother (perhaps something like "mama") came to be seen as standing for the mother. This is consistent with the fact that an extremely high proportion of words for the female parent in both baby talk (Ferguson 1964) and in a corpus of 474 languages (Murdock 1959) have nasal consonants in them.

We argue (MacNeilage & Davis 2004) that following this development a subsequent word for the male parent would have a similar simple structure but would need to contrast phonetically with the word for the female parent. Consistent with this proposal, words for male parent in baby talk (Ferguson 1964) and languages (Murdock 1959) tend to favor oral consonants (e.g., "papa" or "dada").

The word for female parent in this scenario could be regarded as iconic in that it consistently "went with" the female parent as a result of the focus of infant demand on the nearby female parent. However, we argue that that the force towards coining a male parental term that contrasted phonetically with the female term necessarily introduced an element of arbitrariness into the sound-meaning linkage. The conscious realization that arbitrary labels could be attached to concepts, could have started spoken language on its momentous journey with the typical arbitrary relationship between concept and sound pattern that has been so difficult to explain (MacNeilage & Davis 2004).

The baby-talk origins scenario might not seem as plausible as the idea of pantomimes as first words, but it is the only one of the two ideas that is consistent with the *present-day structure* of language, even down to the level of structure of particular lexical items.

ACKNOWLEDGMENT

This paper was prepared with support from research grant No. HD 2773-10 from the U.S. Public Health Service.

Gesture-first, but no gestures?

David McNeill,[a] Bennett Bertenthal,[a] Jonathan Cole,[b] and Shaun Gallagher[c]

[a]*Department of Psychology, The University of Chicago, Chicago, IL 60637;* [b]*Clinical Neurological Sciences, Clinical Neurophysiology, Poole Hospital, Longfleet Road, Poole, BH15 2J, United Kingdom;* [c]*Department of Philosophy, University of Central Florida, Orlando, FL 32816-1352.*
dmcneill@uchicago.edu bbertent@uchicago.edu. jonathan@colefamily.org.uk gallaghr@mail.ucf.edu

Abstract: Although Arbib's extension of the mirror-system hypothesis neatly sidesteps one problem with the "gesture-first" theory of language origins, it overlooks the importance of gestures that occur in current-day human linguistic performance, and this lands it with another problem. We argue that, instead of gesture-first, a system of combined vocalization and gestures would have been a more natural evolutionary unit.

Michael Arbib's extension of the mirror-system hypothesis for explaining the origin of language elegantly sets the stage for further discussion, but we think it overlooks a crucial source of data – the kinds of gestures that actually occur in current human linguistic performance. These data lead us to doubt a basic claim of the "gesture-first" theory, that language started as a gesture language that was gradually supplanted by speech. Arbib has modified this theory with his concept of an expanding spiral, but this new model does not go far enough in representing a speech-gesture system that evolved together.

Classic gesture-first. The enduring popularity of "gesture-first" seems to presuppose that gestures are simple and that as we humans, and language, became more complex, speech evolved and to an extent supplanted gesture, a belief that emerged as part of the Enlightenment quest for the natural state of man and is credited to Condillac, and which has continued since (e.g., Hewes 1973; Armstrong et al. 1995; Corballis 2002). However, contrary to the traditional view, we contend that gesture and language, as they currently exist, belong to a single system of verbalized thinking and communication, and neither can be called the simple twin of the other. It is this *system*, in which both speech and gesture are crucial, that we should be explaining. It makes little sense to ask which part of an unbroken system is "simpler"; a better question is how the parts work together.

In this system, we find synchrony and coexpressiveness – gesture and speech conveying the *same idea unit, at the same time*. Gesture and speech exhibit what Wundt described long ago as the "simultaneous" and "sequential" sides of the sentence (Blumenthal 1970, p. 21) and Saussure, in notes recently discovered, termed "l'essence double du langage" (Harris 2002). Double essence, not enhancement, is the relationship, and we do not see how it could have evolved from the *supplanting* of gestures by speech. In the remainder of this commentary, we summarize three sources of evidence to support this assertion.

1. Consider the attached drawing (Fig. 1). The speaker was describing a cartoon episode in which one character tries to reach another character by climbing up inside a drainpipe. The speaker

Figure 1 (McNeill, et al.). Gesture combining upward movement and interiority. (Computer illustration from a video by Fey Parrill, University of Chicago).

https://doi.org/10.1017/S0140525X05350035 Published online by Cambridge University Press