

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

^aDepartment of Psychology, The University of Chicago, Chicago, IL 60637;

^bClinical Neurological Sciences, Clinical Neurophysiology, Poole Hospital, Longfleet Road, Poole, BH15 2J, United Kingdom; ^cDepartment of Philosophy, University of Central Florida, Orlando, FL 32816-1352.

dmcneill@uchicago.edu bbertent@uchicago.edu
jonathan@cofamily.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).

is saying, “and he goes *up through* the pipe this time,” with the gesture occurring during the boldfaced portion (the illustration captures the moment when the speaker says the vowel of “through”). Coexpressively with “up,” her hand rose upward, and coexpressively with “through,” her fingers spread outward to create an interior space. These took place together and were synchronized with “up through,” the linguistic package that combines the same meanings.

The effect is a uniquely gestural way of packaging meaning – something like “rising hollowness,” which does not exist as a semantic package of English at all. Speech and gesture, at the moment of their synchronization, were coexpressive. The very fact there is shared reference to the character’s climbing up inside the pipe makes clear that it is being represented by the speaker in two ways simultaneously – analytic/combinatoric in speech and global/synthetic in gesture. We suggest it was this very simultaneous combination of opposites that evolution seized upon.

2. When signs and speech do combine in contemporary human performance, they do not synchronize. Kendon (1988) observed sign languages employed by aboriginal Australian women – full languages developed culturally for (rather frequent) speech taboos – which they sometimes combine with speech. The relevant point is that in producing these combinations, speech and sign start out synchronously, but then, as the utterance proceeds, speech outruns the semantically equivalent signs. The speaker stops speaking until the signs catch up and then starts over, only for speech and signs to pull apart again. If, in the evolution of language, there had been a similar doubling up of signs and speech, as the supplanting scenario implies, they too would have been driven apart rather than into synchrony, and for this reason, too, we doubt the replacement hypothesis.

3. The Wundt/Saussure “double essence” of gesture and language appears to be carried by a dedicated thought-hand-language circuit in the brain. This circuit strikes us as a prime candidate for an evolutionary selection at the foundation of language. It implies that the aforementioned *combinations* of speech and gesture were the selected units, not gesture first with speech supplanting or later joining it. We observe this circuit in the unique neurological case of I.W., who lost all proprioception and spatial position sense from the neck down at age 19, and has since taught himself to move using vision and cognition. The thought-language-hand link, located presumably in Broca’s area, ties together language and gesture, and, in I.W., survives and is partly dissociable from instrumental action.

We can address Arbib’s pantomime model by observing the kinds of gestures the dedicated link sustains in I.W.’s performance, in the absence of vision: his gestures are (1) coexpressive and synchronous with speech; (2) not supplemental; and (3) not derivable from pantomime. I.W. is unable to perform instrumental actions without vision but continues to perform speech-synchronized, coexpressive gestures that are virtually indistinguishable from normal (topokinetic accuracy is reduced but morphokinetic accuracy is preserved) (Cole et al. 2002). His gestures without vision, moreover, minimize the one quality that could be derived from pantomime, a so-called “first-person” or “character” viewpoint, in which a gesture replicates an action of a character (cf. McNeill 1992).

More generally, an abundance of evidence demonstrates that spontaneous, speech-synchronized gestures should be counted as part of language (McNeill 1992). Gestures are frequent (accompanying up to 90% of utterances in narrations). They synchronize exactly with coexpressive speech segments, implying that gesture and related linguistic content are coactive in time and jointly convey what is newsworthy in context. Gesture adds cohesion, gluing together potentially temporally separated but thematically related segments of discourse. Speech and gesture develop jointly in children, and decline jointly after brain injury. In contrast to cultural emblems, such as the “O.K.” sign, speech-synchronized gestures occur in all languages, so far as is known. Finally, gestures are not “signs” with an independent linguistic code. Gestures exist only in combination with speech, and are not themselves a coded system.

Arbib’s gesture-first. Arbib’s concept of an expanding spiral may avoid some of the problems of the supplanting mechanism. He speaks of scaffolding and spiral expansion, which appear to mean, in both cases, that one thing is preparing the ground for or propping up further developments of the other thing – speech to gesture, gesture to speech, and so on. This spiral, as now described, brings speech and gesture into temporal alignment (see Fig. 6 in the target article), but also implies two things juxtaposed rather than the evolution of a single “thing” with a double essence. Modification to produce a dialectic of speech and gesture, beyond scaffolding, does not seem impossible. However, the theory is still focused on gestures of the wrong kind for this dialectic – in terms of Kendon’s Continuum (see McNeill 2000 for two versions), signs, emblems, and pantomime. Because it regards all gestures as simplified and meaning-poor, it is difficult to see how the expanding spiral can expand to include the remaining point on the Continuum, “gesticulations” – the kind of speech-synchronized coexpressive gesture illustrated above.

A compromise is that pantomime was the initial protolanguage but was replaced by speech *plus gesture*, leading to the thought-language-hand link that we have described. This hypothesis has the interesting implication that different evolutionary trajectories landed at different points along Kendon’s Continuum. One path led to pantomime, another to coexpressive and speech-synchronized gesticulation, and so on. These different evolutions are reflected today in distinct ways of combining movements with speech. Although we do not question the importance of extending the mirror system hypothesis, we have concerns about a theory that predicts, as far as gesture goes, the evolution of what did not evolve instead of what did.

Meaning and motor actions: Artificial life and behavioral evidence

Domenico Parisi,^a Anna M. Borghi,^b Andrea Di Ferdinando,^c and Giorgio Tsiotas^b

^a*Institute of Cognitive Science and Technologies, National Research Council, Rome 00137, Italy;* ^b*Department of Psychology, University of Bologna, Bologna 40127, Italy;* ^c*Department of Psychology, University of Padua, Padua 35131, Italy.* d.parisi@istc.cnr.it

<http://gral.ip.rm.cnr.it/dparisi/> annamaria.borghi@unibo.it
<http://gral.ip.rm.cnr.it/borghi/> andrea.diferdinando@unipd.it
http://ccnl.psy.unipd.it/di_ferdinando.html giorgio@tsiotas.com

Abstract: Mirror neurons may play a role in representing not only signs but also their meaning. Because actions are the only aspect of behavior that are inter-individually accessible, interpreting meanings in terms of actions might explain how meanings can be shared. Behavioral evidence and artificial life simulations suggest that seeing objects or processing words referring to objects automatically activates motor actions.

Arbib argues that the vocal signs of human language are probably evolved from the gestural signs of some protolanguage, and this might explain why the production of vocal signs in the human brain is controlled by Broca’s area – which corresponds to area V5 in monkeys’ brain – which controls manual actions. The discovery of neurons in both areas that are activated both when a manual action is executed and when it is observed in others (mirror neurons) reinforces this interpretation, because language is based on what Arbib calls the parity requirement, according to which what counts for the speaker must count approximately the same for the hearer.

However, language is not only signs but is signs plus the meaning of signs. Mirror neurons tend to be invoked to explain the production of linguistic signs but they may also play an important role in the representation of the meaning of those signs. If meanings are interpreted as categories of entities in the environment, one can argue that these categories are represented in the brain in