diovisual property of these mirror neurons puts them in position to form a special kind of abstraction. Many of the neurons respond equally well to the sight of an action and to the sound associated with an action (Keysers et al. 2003). This means that they are representing an action not only regardless of who performs it, but also regardless of the modality through which it is perceived. The multimodality of this kind of representation may have been an important step towards the use of the motor system in symbolic language. Performed and observed actions can be associated with both sounds and sights. This makes the motor cortex a prime candidate as a potential locus for the development of multimodal (or amodal) representations, which are so important to language.

Support for this view comes from an fMRI study we recently conducted on audiovisual interactions in the perception of actions (Kaplan & Iacoboni, submitted). When subjects saw and heard an action (i.e., tearing paper) simultaneously, there was greater activity in the left ventral premotor cortex compared with control conditions in which they only saw or only heard the action. This cross-modal interaction did not happen with a non-action control stimulus (i.e., a square moving while a sound was played), suggesting that the premotor cortex is sensitive to the conjunction of visual and auditory representations of an action. Again, it may be this capacity for conjunctive representations that led to true symbolic capability.

Further support for the role of the auditory responsiveness of motor neurons in language evolution comes from transcranial magnetic stimulation (TMS) studies on motor facilitation in the two cerebral hemispheres in response to the sight or the sound of an action. Motor activation to the sight of an action is typically bilateral, albeit slightly larger in the left hemisphere in right-handers (Aziz-Zadeh et al. 2002). Action sounds, in contrast, activate the motor cortex only in the left hemisphere, the cerebral hemisphere dominant for language (Aziz-Zadeh et al. 2004). Since there is no evidence of lateralized auditory responses of mirror neurons in the monkey, the lateralization for action sounds observed in the TMS study and the lateralization of cross-modal interactions in the ventral premotor cortex seem to be related to evolutionary processes that made human brain functions such as language lateralized to the left hemisphere.

A more central role of auditory properties of mirror neurons in language evolution makes also the transition from manual gestures to mouth-based communication (speech) easier to account for. Recent fMRI data suggest that the human premotor cortex seems able to map some kind of articulatory representation onto almost any acoustic input (Schubotz & von Cramon 2003). A multi-sensory representation of action provided by mirror neurons responding also to action sounds may have more easily evolved in articulatory representation of the sounds associated with manual actions.

In summary, it may be the premotor cortex's unique position of having both cross-modal and cross-agent information that allowed it to support language. The auditory properties of mirror neurons may have been a facilitator rather than a by-product of language evolution.

Pragmatics, prosody, and evolution: Language is more than a symbolic system

Boris Kotchoubey

Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, 72074 Tübingen, Germany.
boris.kotchoubey@uni-tuebingen.de
http://www.uni-tuebingen.de/medizinischepsychologie/stuff/

Abstract: The model presented in the target article is biased towards a cognitive-symbolic understanding of language, thus ignoring its other important aspects. Possible relationships of this cognitive-symbolic subsystem to pragmatics and prosody of language are discussed in the first part of the commentary. In the second part, the issue of a purely social versus biological mechanisms for transition from protolanguage to properly language is considered

1. Arbib's conception of language, summarised in LA1 to LA4, is concentrated upon its cognitive components and the cognitive abilities that both underlie and are based on verbal communication. Although semantics and syntax are the only components of the language in highly intelligent speaking robots, human languages also include expressive components such as intonation and gesticulation. Particularly, prosody subserves two important functions of emotional expression (affective prosody) and of clarification of the content's meaning (linguistic prosody, such as distinguishing between an assertion and a question) (Bostanov & Kotchoubey 2004; Seddoh 2002). Neuropsychological and neuroimaging data converge in demonstrating that both linguistic and affective prosodic information is processed mainly in the right temporal lobe (Ross 1981), in contrast to semantics and syntax, which are processed in the left temporal lobe. Affective prosody is strikingly similar in humans and other primates, so that human subjects having no previous experience with monkeys correctly identify the emotional content of their screams (Linnankoski et al. 1994).

It is therefore tempting to represent the system of language as entailing two virtually additive subsystems. The left hemispheric subsystem develops on the basis of the mirror system of apes in an indirect way depicted in the target article, and subserves the cognitive-symbolic function of language, its referential network, and syntactic design. The right hemispheric subsystem, in contrast, is a direct successor of monkeys' vocalisation mechanisms and gives our language its intonational colour and expressive power (Scherer 1986).

This view would ignore, however, the possibly most important aspect of language: its pragmatics. Except for some scientific discussions, which did not play any important role before 2,500 years ago (and even after this point their role should not be overestimated), communication is directed to move somebody to do something. Communication is only a means, whereas the goal is co-operation.[1] The pragmatic function of language goes beyond the mere referential semantics and mere expression of one's own state: It links together verbal and non-verbal, symbolic and non-symbolic components of language because it relates us, over all conventional symbols (words), to some, perhaps very remote, non-conventional basis. Likewise, affective prosody is not symbolic and conventional; it is a part of emotion itself. This pragmatic view makes it very difficult to imagine a certain moment in the evolution of language when its left- and right-hemispheric components met together; rather, they should have been together from the very beginning.

Some recent neuropsychological data point in the same direction. Although the right temporal lobe is critical for recognition of prosody (Adolphs et al. 2002), prosodic aspects of language are also severely impaired in patients with lesions to orbitofrontal cortex (Hornak et al. 2003) and the corpus callosum (Friederici et al. 2003). All this makes the simple additive model (i.e., the ancient prosodic subsystem is simply added to the newly developed cognitive subsystem) implausible. Rather, a theory is needed that would describe the development of language in mutual interaction of its different aspects.

2. Arbib suggests that the development of language from protolanguage was a social rather than biological process. The only mechanism of such social progress he describes in section 7 is the unexpected and unpredictable linguistic inventions made by numerous but anonymous genii, those inventions being then seized upon and employed by other people. I agree that no other social mechanism can be thought of, because otherwise social systems are usually conservative and favour hampering, rather than promoting, development (e.g., Janis 1982). Surely, this putative process of social inventions is familiar: somebody has a good idea, others learn about it, after a period of resistance they become accustomed to it and see its advantages, and soon the whole social

https://doi.org/10.1017/S0140525X05340039 Published online by Cambridge University Press

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