For these reasons, we suggested that this frontal auditory/motor domain may belong to, or be the precursor of, a vocalization mirror system similar to the mirror system for grasping, which in hominids participated in vocal imitative behavior, allowing them to compare heard vocalizations with their own productions (Bosman et al. 2004; Jürgens 2003). All it would take to develop this system into a complex, voluntary vocalizing system might be a refinement of the respective circuits and increasing cortico-bulbar control. In this line, evidence indicates a phylogenetic trend from nonhuman primates to humans towards increasing cortical control of the tongue, which may be related to the superior role the tongue plays in speech (Jürgens & Alipour 2002).

In parallel to this evidence, a very recent fMRI study has demonstrated that in humans, listening to speech activates a superior portion of the ventral premotor cortex that largely overlaps with a speech-production motor area (Wilson et al. 2004). This evidence suggests the existence of a human vocalization mirror system, perhaps derived from the regions in the monkey described above. In consequence, we think that a more parsimonious hypothesis could be that instead of a serial dependence of vocal communication upon gestural communication, both coevolved to a large extent; that is, both developed their own circuitry in parallel, with a high degree of interaction between the two systems (Izumi & Kojima 2004).

Against these arguments, it has been claimed that in nonhuman primates, cortical control over hand movements is stronger than control of vocalizations, which partly explains why apes can be taught sign language and not vocal communication (Corballis 2003a). However, in our view this does not imply that gestural communication must be ancestral to vocal communication. The same or even more behavioral flexibility (including combinatorial abilities) than that observed in hand coordination, may have developed in vocal communication by elaborating on preexisting vocal circuits. A similar situation may be observed in the elephant's trunk: the neural machinery controlling the trunk probably developed on its own, without the necessity of borrowing a coordination system from other motor devices (Pinker 1995). In addition, the presumed ancestral signing stage remains highly speculative, there being still no evidence for it. Summarizing, since in monkeys and apes most communication is vocal, and given that there is an incipient prefrontal control for vocalizations in them, we see no necessity to propose a stage of gestural communication preceding "protospeech."

Finally, we would like to comment on the contrast previously made by Arbib and Bota (2003), which we think may be misleading, between their theory being "prospective" (finding what is in the monkey – hand coordination – which may have served as a substrate for human language), and our theory (Aboitiz & García 1997) being "retrospective" (looking at what is in the human brain – working memory – and tracking it back to the monkey brain). Aboitiz and García (1997) followed standard phylogenetic methodology: first, the study identified in the monkey the networks that can be homologous to the language-related neural networks; second, it asked about the functions of these networks in the monkey and in the human, one of which is working memory.

A good analogy for this strategy comes from the evolution of the eye (Dawkins 1996): Although image formation is a highly derived characteristic, there are more basic functions such as photoreception, which are central to vision and shared by other species whose visual organs lack image-forming properties; these functions permit us to track the phylogenetic ancestry of the eyes. Likewise, Aboitiz and García (1997) point to a function (working memory) that is present in both the human and the monkey and participates in language processing (Aboitiz et al., in press; Smith & Jonides 1998). On the other hand, although hand coordination networks are present in both species, at this point there is no evidence for the involvement of the hand control system in human linguistic processing.

ACKNOWLEDGMENTS
Part of the work referred to in this commentary was financed by the Millennium Center for Integrative Neuroscience. C. Bosman is supported by MECESUP PUC0005.

Action planning supplements mirror systems in language evolution

Bruce Bridgeman
Department of Psychology, Social Sciences 2, University of California, Santa Cruz, Santa Cruz, CA 95064. **bruceb@ucsc.edu**
http://psych.ucsc.edu/Faculty/bBridge.shtml

Abstract: Mirror systems must be supplemented by a planning capability to allow language to evolve. A capability for creating, storing, and executing plans for sequences of actions, having evolved in primates, was applied to sequences of communicatory acts. Language could exploit this already-existing capability. Further steps in language evolution may parallel steps seen in the development of modern children.

Because the functional basis for language capability lies in the brain, it is sensible to look to brain evolution for insight into the evolution of language. Though the recently discovered mirror system in primates offers possibilities for the evolution of capabilities necessary for language, it is not enough to do the whole job. Indeed, the well-developed mirror system of monkeys in the absence of language shows that something more is needed, as Arbib points out. In emphasizing the mirror neuron system, a here-and-now system, Arbib makes a convincing case that mirror neurons are important in language evolution. A second need is for hierarchical structure rather than mere sequencing (target article, sect. 7, para. 13). This commentary will elaborate on that need and how it is met.

A key power of language is the use of sequences of symbols in a grammatical system. For the ability to handle sequences, evolution of primate planning mechanisms is essential. Complementary to the mirror-neuron evolution story is the increasing ability of primates to plan sequences of actions, for instance in preparing and using tools. Actions must be planned in the brain before the sequence starts, and must be executed in a particular order to achieve success. The organization is hierarchical, with smaller tasks embedded in larger ones. The lateral prefrontal cortex is probably the location of the machinery that produces, stores, and executes such plans. As planning abilities improved over the course of primate evolution, the planning of sequences of actions loomed ever greater in importance.

In this conception, a critical event in the evolution of language was the use of this growing capability for planning to generate not sequences of actions, but sequences of words (Bridgeman 1992). This idea addresses two of the central puzzles of language evolution – first, how such a complex capability could evolve in such a short time, and second, how it could evolve in small steps, each useful immediately. The solution to the first problem is that language is a new capability made mostly of old neurological parts, among them the mirror system and the planning capability.

To examine the second problem, the small steps, we can look to human development for hints about how the evolution of language may have proceeded, to the genetic remnants of earlier adaptations that remain in modern humans. The importance of gesture is clear from ontogeny as well as neurology, as most infants achieve a well-developed gestural communication before the first word. The gestures, although they eclipse the stereotyped call systems of other animals, remain single communications fixed in the here-and-now. The first words occur in combination with gesture and context to create useful communications with minimal verbal content.

Arbib's suggestion (sect. 2, para. 2) – that single utterances of *Homo erectus* and early *Homo sapiens* could convey complex meanings that modern languages achieve only with longish phrases – is unlikely to be accurate. Arbib's comparison to monkey calls demonstrates this; most of them can be paraphrased in one or two words; "leopard," "I'm angry," and so on. Similarly, an infant's first words are at the monkey-call level of generalization, not the whole sentence in a word that Arbib imagines. Arbib's suggestion would require that super-words and the capacity to de-

https://doi.org/10.1017/S0140525X0526003X Published online by Cambridge University Press

velop and use them evolved, then disappeared again in favor of the more specific words that characterize all existing languages. All this would have had to occur before speaking hominids gave rise to the present population, because the generality of words is about the same in all languages and therefore probably constitutes a "universal" of language, that is, a species-specific and possibly a part of our biological language equipment.

One-word phrases address one of the paradoxes of language evolution: in order to create a selective pressure for evolution of better capability in using grammar, there must be a preexisting, culturally defined lexicon with which the grammar can be built. Many of the words used in modern languages could appear in this way, but others, especially modifiers such as tense markers, cannot. At this stage, words name things. The thing can be an object (later, noun), an action (verb), or a property (adjective/adverb). Again, the paradox is the same: that such modifiers would have to exist already before a complex grammar could develop.

How could the sorts of words that cannot be used alone get invented? Again we have evidence from the development of language in children. True, a child's first words are single "holophrase" utterances, often comprehensible only in a context. But next comes a two-word slot grammar, the same all over the world regardless of the structure of the parent language. This suggests a biologically prepared mechanism (reviewed in Bridgeman 2003, Ch. 7). Culturally, a large lexicon could develop at this stage, more complex than one-word phrases could support, making possible and useful the further development of grammar.

Though the slot grammar of toddlers is different from that of the child's eventual language, it has several properties that make it useful for developing structure in a lexicon. Single-word utterances need not differentiate parts of speech, since there is no grammar. Words such as "sour" and "fruit" would be parallel – descriptions of some property of the world. Only when combined with another word must they be differentiated. Most of the utterances of the slot grammar consist of a noun and a modifier, either an adjective or a verb, that qualifies the context of the noun.

A "language" such as this is severely limited. We can imagine some group of *Homo erectus* sitting around their fire after a hard day of hunting and gathering. Someone announces, "Lake cold." Another replies, "Fishing good." The results seem almost comical to us, but such terms would be tremendously more useful than no language at all, because they allow the huge advantage that humans have over other living primates – to allow the experience of one individual to increase the knowledge of another. Once this level of communication is achieved, the selective pressure would be tremendous to develop all the power and subtlety of modern language.

Sign languages are problematic for a gestural origins theory of language evolution

Karen Emmorey

Laboratory for Cognitive Neuroscience, The Salk Institute for Biological Studies, La Jolla, CA 92037. **emmorey@salk.edu**
http://www-psy.ucsd.edu:80/~kemmorey

Abstract: Sign languages exhibit all the complexities and evolutionary advantages of spoken languages. Consequently, sign languages are problematic for a theory of language evolution that assumes a gestural origin. There are no compelling arguments why the expanding spiral between protosign and protospeech proposed by Arbib would not have resulted in the evolutionary dominance of sign over speech.

At first glance, the existence of modern sign languages provides support for Arbib's hypothesis that there was an early stage in the evolution of language in which communication was predominantly gestural. Modern sign languages offer insight into how pantomimic communication might have evolved into a more language-like system (i.e., protosign). Diachronic linguistic analyses have traced grammaticalization pathways in American Sign Language (ASL) that originate with gesture (Janzen & Shaffer 2002). For example, grammatical markers of modality in ASL (e.g., "can," "must") are derived from lexical signs ("strong," "owe"), and these lexical signs are in turn derived from nonlinguistic communicative gestures (clenching the fists and flexing muscles to indicate strength and a deictic pointing gesture indicating monetary debt). Investigations of newly emerging signed languages are also uncovering patterns of conventionalization and grammaticalization that originate in pantomimic and communicative gestures (e.g., Kegl et al. 1999). Of course, these are modern sign languages acquired and created by modern human brains, but the evidence indicates that communicative gestures can evolve into language.

Arbib reasonably proposes that the transition from gesture to speech was not abrupt, and he suggests that protosign and protospeech developed in an expanding spiral until protospeech became dominant for most people. However, there is no evidence that protosign ever became dominant for any subset of people – except for those born deaf. The only modern communities in which a signed language is dominant have deaf members for whom a spoken language cannot be acquired naturally. No known community of hearing people (without deaf members) uses a signed language as the primary language. Hence, a community of *deaf* people appears to be a prerequisite for the emergence and maintenance of a sign language. Although it is possible that a sign language (and its deaf community) has existed for 6,000 years (the divergence date for Indo-European spoken languages), the earliest known sign language can be tentatively traced back only 500 years to the use of Turkish Sign Language at the Ottoman court (Zeshan 2003).

The fact that signed languages appear to be relatively new languages does not mean that they are somehow inferior to spoken languages. Signed languages are just as complex, just as efficient, and just as useful as spoken languages. Signed languages easily express abstract concepts, are acquired similarly by children, and are processed by the same neural systems within the left hemisphere (see Emmorey 2002 for review). Thus, in principle, there is no linguistic reason why the expanding spiral between protosign and protospeech could not have resulted in the evolutionary dominance of sign over speech. A gestural-origins theory must explain why speech evolved at all, particularly when choking to death is a potential by-product of speech evolution due to the repositioning of the larynx.

Corballis (2002) presents several specific hypotheses why speech might have won out over gesture, but none are satisfactory (at least to my mind). Corballis suggests that speech may have an advantage because more arbitrary symbols are used, but sign languages also consist of arbitrary symbols, and there is no evidence that the iconicity of some signs limits expression or processing. The problem of signing in the dark is another oft-cited disadvantage for sign language. However, early signers/gesturers could sign in moonlight or firelight, and a tactile version of sign language could even be used if it were pitch black (i.e., gestures/signs are felt). Furthermore, speech has the disadvantage of attracting predators with sound at night or alerting prey during a hunt. Corballis argues that speech would allow for communication simultaneously with manual activities, such as tool construction or demonstration. However, signers routinely sign with one hand, while the other hand holds or manipulates an object (e.g., turning the steering wheel while driving and signing to a passenger). It is true that operation of a tool that requires two hands would necessitate serial manual activity, interspersing gesturing with object manipulation. But no deaths have occurred from serial manual activity, unlike the deaths that occur as a result of choking.

Everyone agrees that the emergence of language had clear and compelling evolutionary advantages. Presumably, it was these advantages that outweighed the dangerous change in the vocal tract that allowed for human speech but increased the likelihood of choking. If communicative pantomime and protosign preceded

https://doi.org/10.1017/S0140525X0526003X Published online by Cambridge University Press