From Monkey-Like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics

Article *in* Behavioral and Brain Sciences · May 2005

DOI: 10.1017/S0140525X05000038 · Source: PubMed

Michael A Arbib
University of California, San Diego

BEHAVIORAL AND BRAIN SCIENCES (2005) **28,** 105–167
Printed in the United States of America

From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics

Michael A. Arbib
Computer Science Department, Neuroscience Program, and USC Brain Project, University of Southern California, Los Angeles, CA 90089-2520
arbib@pollux.usc.edu http://www-hbp.usc.edu/

Abstract: The article analyzes the neural and functional grounding of language skills as well as their emergence in hominid evolution, hypothesizing stages leading from abilities known to exist in monkeys and apes and presumed to exist in our hominid ancestors right through to modern spoken and signed languages. The starting point is the observation that both premotor area F5 in monkeys and Broca's area in humans contain a "mirror system" active for both execution and observation of manual actions, and that F5 and Broca's area are homologous brain regions. This grounded the mirror system hypothesis of Rizzolatti and Arbib (1998) which offers the mirror system for grasping as a key neural "missing link" between the abilities of our nonhuman ancestors of 20 million years ago and modern human language, with manual gestures rather than a system for vocal communication providing the initial seed for this evolutionary process. The present article, however, goes "beyond the mirror" to offer hypotheses on evolutionary changes within and outside the mirror systems which may have occurred to equip *Homo sapiens* with a language-ready brain. Crucial to the early stages of this progression is the mirror system for grasping and its extension to permit imitation. Imitation is seen as evolving via a so-called simple system such as that found in chimpanzees (which allows imitation of complex "object-oriented" sequences but only as the result of extensive practice) to a so-called complex system found in humans (which allows rapid imitation even of complex sequences, under appropriate conditions) which supports pantomime. This is hypothesized to have provided the substrate for the development of protosign, a combinatorially open repertoire of manual gestures, which then provides the scaffolding for the emergence of protospeech (which thus owes little to nonhuman vocalizations), with protosign and protospeech then developing in an expanding spiral. It is argued that these stages involve biological evolution of both brain and body. By contrast, it is argued that the progression from protosign and protospeech to languages with full-blown syntax and compositional semantics was a historical phenomenon in the development of *Homo sapiens*, involving few if any further biological changes.

Key words: gestures; hominids; language evolution; mirror system; neurolinguistics; primates; protolanguage; sign language; speech; vocalization

1. Action-oriented neurolinguistics and the mirror system hypothesis

1.1. Evolving the language-ready brain

Two definitions:

1. A *protolanguage* is a system of utterances used by a particular hominid species (possibly including *Homo sapiens*) which we would recognize as a precursor to human language (if only the data were available!), but which is not itself a human language in the modern sense.[1]

2. An infant (of any species) has a *language-ready* brain if it can acquire a full human language when raised in an environment in which the language is used in interaction with the child.

Does the language readiness of human brains require that the richness of syntax and semantics be encoded in the genome, or is language one of those feats – from writing history to building cities to using computers – that played no role in biological evolution but rested on historical developments that created societies that could develop and transmit these skills? My hypothesis is that:

> Language readiness evolved as a multimodal manual/facial/vocal system with protosign (manual-based protolanguage) providing the scaffolding for protospeech (vocal-based protolanguage) to provide "neural critical mass" to allow language to emerge from protolanguage as a result of cultural innovations within the history of *Homo sapiens*.[2]

The theory summarized here makes it understandable why it is as easy for a deaf child to learn a signed language as it is for a hearing child to learn a spoken language.

MICHAEL ANTHONY ARBIB was born in England, grew up in Australia, and received his Ph.D. in Mathematics from MIT. After five years at Stanford, he became chairman of Computer and Information Science at the University of Massachusetts, Amherst, in 1970. He moved to the University of Southern California in 1986, where he is Professor of Computer Science, Neuroscience, Biomedical Engineering, Electrical Engineering, and Psychology. The author or editor of 38 books, Arbib recently co-edited *Who Needs Emotions? The Brain Meets the Robot* (Oxford University Press) with Jean-Marc Fellous. His current research focuses on brain mechanisms of visuomotor behavior, on neuroinformatics, and on the evolution of language.

0140-525X/05 $12.50

1.2. The mirror system hypothesis

Humans, chimps and monkeys share a general physical form and a degree of manual dexterity, but their brains, bodies, and behaviors differ. Moreover, humans can and normally do acquire language, and monkeys and chimps cannot – though chimps and bonobos can be trained to acquire a form of communication that approximates the complexity of the utterances of a 2-year-old human infant. The approach offered here to the evolution of brain mechanisms that support language is anchored in two observations: (1) The system of the monkey brain for visuomotor control of hand movements for grasping has its premotor outpost in an area called F5 which contains a set of neurons, called *mirror neurons*, each of which is active not only when the monkey executes a specific grasp but also when the monkey observes a human or other monkey execute a more or less similar grasp (Rizzolatti et al. 1996a). Thus F5 in monkey contains a *mirror system for grasping* which employs a common neural code for *executed* and *observed* manual actions (sect. 3.2 provides more details). (2) The region of the human brain homologous to F5 is part of Broca's area, traditionally thought of as a speech area but which has been shown by brain imaging studies to be active when humans both execute and observe grasps.

These findings led to the mirror system hypothesis (Arbib & Rizzolatti 1997; Rizzolatti & Arbib 1998, henceforth R&A):

> The *parity requirement* for language in humans – that what counts for the speaker must count approximately the same for the hearer[3] – is met because Broca's area evolved atop the mirror system for grasping, with its capacity to generate and recognize a set of actions.

One of the contributions of this paper will be to stress that the F5 mirror neurons in the monkey are linked to regions of parietal and temporal cortex, and then argue that the evolutionary changes that "lifted" the F5 homologue of the common ancestor of human and monkey to yield the human Broca's area also "lifted" the other regions to yield Wernicke's area and other areas that support language in the human brain.

Many critics have dismissed the mirror system hypothesis, stating correctly that monkeys do not have language and so the mere possession of a mirror system for grasping cannot suffice for language. But the key phrase here is "evolved atop" – and Rizzolatti and Arbib (1998) discuss explicitly how changes in the primate brain might have adapted the use of the hands to support pantomime (intended communication) as well as praxis, and then outlined how further evolutionary changes could support language. The hypothesis provides a neurological basis for the oft-repeated claim that hominids had a (proto)language based primarily on manual gestures before they had a (proto)language based primarily on vocal gestures (e.g., Armstrong et al. 1995; Hewes 1973; Kimura 1993; Stokoe 2001).[4] It could be tempting to hypothesize that certain species-specific vocalizations of monkeys (such as the snake and leopard calls of vervet monkeys) provided the basis for the evolution of human speech, since both are in the vocal domain. However, these primate vocalizations appear to be related to non-cortical regions as well as the anterior cingulate cortex (see, e.g., Jürgens 1997) rather than F5, the homologue of Broca's area. I think it likely (though empirical data are sadly lacking) that the primate cortex contains a mirror system for such species-specific vocalizations, and that a related mirror system persists in humans, but I suggest that it is a complement to, rather than an integral part of, the speech system that includes Broca's area in humans.

The mirror system hypothesis claims that a *specific* mirror system – the primate mirror system for grasping – evolved into a key component of the mechanisms that render the human brain language-ready. It is this specificity that will allow us to explain below why language is multimodal, its evolution being based on the execution and observation of hand movements. There is no claim that mirroring or imitation is limited to primates. It is likely that an analogue of mirror systems exists in other mammals, especially those with a rich and flexible social organization. Moreover, the evolution of the imitation system for learning songs by male songbirds is divergent from mammalian evolution, but for the neuroscientist there are intriguing challenges in plotting the similarities and differences in the neural mechanisms underlying human language and birdsong (Doupe & Kuhl 1999).[5]

The monkey mirror system for grasping is presumed to allow other monkeys to understand praxic actions and use this understanding as a basis for cooperation, averting a threat, and so on. One might say that this is *implicitly* communicative, as a side effect of conducting an action for non-communicative goals. Similarly, the monkey's orofacial gestures register emotional state, and primate vocalizations can also communicate something of the current priorities of the monkey, but to a first order this might be called "involuntary communication"[6] – these "devices" evolved to signal certain aspects of the monkey's current internal state or situation either through its observable actions or through a fixed species-specific repertoire of facial and vocal gestures. I will develop the hypothesis that the mirror system made possible (but in no sense guaranteed) the evolution of the displacement of hand movements from praxis to gestures that can be controlled "voluntarily."

It is important to be quite clear as to what the mirror system hypothesis does *not* say.

1. It does not say that having a mirror system is equivalent to having language. Monkeys have mirror systems but do not have language, and I expect that many species have mirror systems for varied socially relevant behaviors.
2. Having a mirror system for grasping is not in itself sufficient for the copying of actions. It is one thing to recognize an action using the mirror system; it is another thing to use that representation as a basis for repeating the action. Hence, *further evolution of the brain was required for the mirror system for grasping to become an imitation system for grasping*.
3. It does not say that language evolution can be studied in isolation from cognitive evolution more generally.

Arbib (2002) modified and developed the R&A argument to hypothesize seven stages in the evolution of language, with imitation grounding two of the stages.[7] The first three stages are pre-hominid:

S1: Grasping.

S2: A mirror system for grasping shared with the common ancestor of human and monkey.

S3: A simple imitation system for object-directed grasping through much-repeated exposure. This is shared with the common ancestor of human and chimpanzee.

The next three stages then distinguish the hominid line from that of the great apes:

S4: A complex imitation system for grasping – the ability to recognize another's performance as a set of familiar actions and then repeat them, or to recognize that such a performance combines novel actions which can be approximated by variants of actions already in the repertoire.[8]

S5: *Protosign*, a manual-based communication system, breaking through the fixed repertoire of primate vocalizations to yield an open repertoire.

S6: *Protospeech*, resulting from the ability of control mechanisms evolved for protosign coming to control the vocal apparatus with increasing flexibility.[9]

The final stage is claimed (controversially!) to involve little if any biological evolution but instead to result from cultural evolution (historical change) in *Homo sapiens*:

S7: *Language*, the change from action-object frames to verb-argument structures to syntax and semantics; the coevolution of cognitive and linguistic complexity.

The Mirror System Hypothesis is simply the assertion that the mechanisms that get us to the role of Broca's area in language depend in a crucial way on the mechanisms established in stage S2. The above seven stages provide just one set of hypotheses on how this dependence may have arisen. The task of this paper is to re-examine this progression, responding to critiques by amplifying the supporting argument in some cases and tweaking the account in others. I believe that the overall framework is robust, but there are many details to be worked out and a continuing stream of new and relevant data and modeling to be taken into account.

The claim for the crucial role of manual communication in language evolution remains controversial. MacNeilage (1998; MacNeilage & Davis, in press b), for example, has argued that language evolved directly as speech. (A companion paper [Arbib 2005] details why I reject MacNeilage's argument. The basic point is to distinguish the evolution of the ability to use gestures that convey meaning from the evolution of syllabification as a way to structure vocal gestures.)

A note to commentators: The arguments for stages S1 through S6 can and should be evaluated quite independently of the claim that the transition to language was cultural rather than biological.

The neurolinguistic approach offered here is part of a performance approach which explicitly analyzes both perception and production (Fig. 1). For production, we have much we could possibly talk about which is represented as cognitive structures (cognitive form; schema assemblages) from which some aspects are selected for possible expression. Further selection and transformation yields semantic structures (hierarchical constituents expressing objects, actions and relationships) which constitute a semantic form that is enriched by linkage to schemas for perceiving and acting upon the world (Arbib 2003; Rolls & Arbib 2003). Finally, the ideas in the semantic form must be expressed in words whose markings and ordering are expressed in phonological form – which may include a wide range of ordered expressive gestures, whether manual, orofacial, or vocal. For perception, the received sentence must be interpreted semantically, with the result updating the "hearer's" cognitive structures. For example, perception of a visual scene may reveal "Who is doing what and to whom/which" as part of a nonlinguistic *action-object frame* in cognitive form. By contrast, the *verb-argument structure* is an overt linguistic representation in semantic form – in modern human languages, generally the action is named by a verb and the objects are named by nouns or noun phrases (see sect. 7). A production grammar for a language is then a specific mechanism (whether explicit or implicit) for converting verb-argument structures into strings of words (and hierarchical compounds of verb-argument structures into complex sentences), and vice versa for perception.

In the brain there may be no single grammar serving both production and perception, but rather, a "direct grammar" for production and an "inverse grammar" for perception. Jackendoff (2002) offers a competence theory with a much closer connection with theories of processing than has been common in generative linguistics and suggests (his sect. 9.3) strategies for a two-way dialogue between competence and performance theories. Jackendoff's approach to competence appears to be promising in this regard because it at-

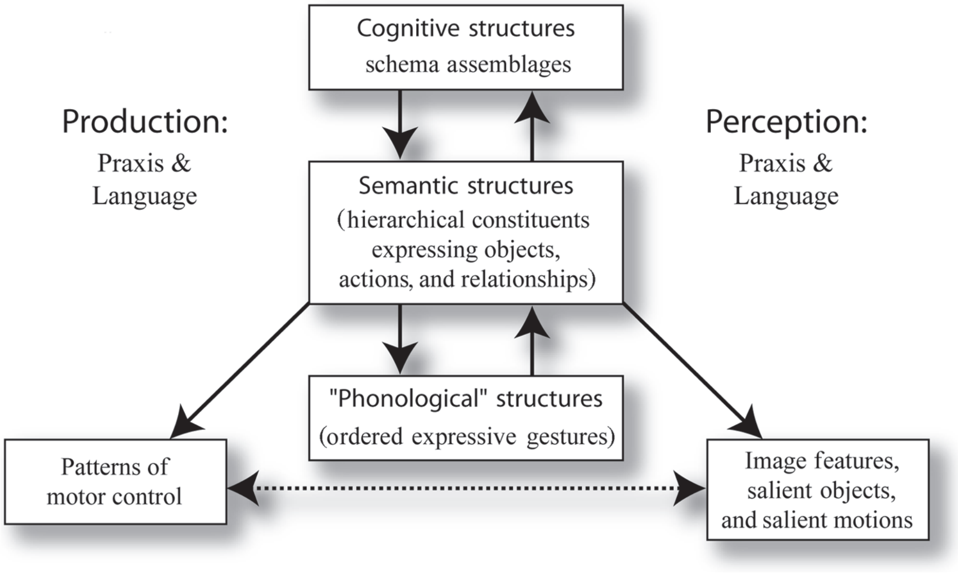


Figure 1. A performance view of the production and perception of language.

tends to the interaction of, for example, phonological, syntactic, and semantic representations. There is much, too, to be learned from a variety of approaches to cognitive grammar which relates cognitive form to syntactic structure (see, e.g., Heine 1997; Langacker 1987; 1991; Talmy 2000).

The next section provides a set of criteria for language readiness and further criteria for what must be added to yield language. It concludes (sect. 2.3) with an outline of the argument as it develops in the last six sections of the paper.

2. Language, protolanguage, and language readiness

I earlier defined a protolanguage as any system of utterances which served as a precursor to human language in the modern sense and hypothesized that the first *Homo sapiens* had protolanguage and a "language-ready brain" but did not have language.

Contra Bickerton (see Note 1), I will argue in section 7 that the prelanguage of *Homo erectus* and early *Homo sapiens* was composed mainly of "unitary utterances" that symbolized frequently occurring situations (in a general sense) without being decomposable into distinct words denoting components of the situation or their relationships. Words as we know them then co-evolved culturally with syntax through fractionation. In this view, many ways of expressing relationships that we now take for granted as part of language were the discovery of *Homo sapiens*; for example, adjectives and the fractionation of nouns from verbs may be "post-biological" in origin.

2.1. Criteria for language readiness

Here are properties hypothesized to support protolanguage:

LR1. *Complex imitation*: The ability to recognize another's performance as a set of familiar movements and then repeat them, but also to recognize that such a performance combines novel actions that can be approximated by (i.e., more or less crudely be imitated by) variants of actions already in the repertoire.[10]

The idea is that this capacity – distinct from the simple imitation system for object-directed grasping through much repeated exposure which is shared with chimpanzees – is necessary to support properties LR2 and LR3, including the idea that symbols are potentially arbitrary rather than innate:

LR2. *Symbolization*: The ability to associate symbols with an open class of episodes, objects, or actions.

At first, these symbols may have been unitary utterances, rather than words in the modern sense, and they may have been based on manual and facial gestures rather than being vocalized.

LR3. *Parity (mirror property)*: What counts for the speaker (or producer) must count for the listener (or receiver).

This extends Property LR2 by ensuring that symbols can be shared, and thus is bound up with LR4.

LR4. *Intended communication*: Communication is intended by the utterer to have a particular effect on the recipient rather than being involuntary or a side effect of praxis.

The remainder are more general properties, delimiting cognitive capabilities that underlie a number of the ideas which eventually find their expression in language:

LR5. *From hierarchical structuring to temporal ordering*: Perceiving that objects and actions have subparts; finding the appropriate timing of actions to achieve goals in relation to those hierarchically structured objects.

A basic property of language – translating a hierarchical conceptual structure into a temporally ordered structure of actions – is in fact not unique to language but is apparent whenever an animal takes in the nature of a visual scene and produces appropriate behavior. Animals possess subtle mechanisms of action-oriented perception with no necessary link to the ability to communicate about these components and their relationships. To have such structures does not entail the ability to communicate by using words or articulatory gestures (whether signed or vocalized) in a way that reflects these structures.

Hauser et al. (2002) assert that the faculty of language in the narrow sense (FLN) includes only recursion and is the one uniquely human component of the faculty of language. However, the flow diagram given by Byrne (2003) shows that the processing used by a mountain gorilla when preparing bundles of nettle leaves to eat is clearly recursive. Gorillas (like many other species, and not only mammals) have the working memory to refer their next action not only to sensory data but also to the state of execution of some current plan. Hence, when we refer to the monkey's grasping and ability to recognize similar grasps in others, it is a mistake to treat the individual grasps in isolation – the F5 system is part of a larger system that can direct those grasps as part of a recursively structured plan.

Let me simply list the next two properties here, and then expand upon them in the next section:

LR6. *Beyond the here-and-now 1*: The ability to recall past events or imagine future ones.

LR7. *Paedomorphy and sociality*: Paedomorphy is the prolonged period of infant dependency which is especially pronounced in humans; this combines with social structures for caregiving to provide the conditions for complex social learning.

Where Deacon (1997) makes symbolization central to his account of the coevolution of language and the human brain, the present account will stress the parity property LR3, since it underlies the sharing of meaning, and the capacity for complex imitation. I will also argue that only protolanguage co-evolved with the brain, and that the full development of linguistic complexity was a cultural/historical process that required little or no further change from the brains of early *Homo sapiens*.

Later sections will place LR1 through LR7 in an evolutionary context (see sect. 2.3 for a summary), showing how the coupling of complex imitation to complex communication creates a language-ready brain.

2.2. Criteria for language

I next present four criteria for what must be added to the brain's capabilities for the parity, hierarchical structuring, and temporal ordering of language readiness to yield *language*. Nothing in this list rests on the medium of exchange of the language, applying to spoken language, sign language, or written language, for example. My claim is that a brain that can support properties LR1 through LR7 above can support properties LA1 through LA4 below – as long as its "owner" matures in a society that possesses language in the sense so defined and nurtures the child to acquire it. In other words, I claim that the mechanisms that make LR1 through LR7 possible are supported by the genetic encoding of brain and body and the consequent space of possible social interactions, but that the genome has no additional structures specific to LA1 through LA4. In particular, the genome does not have special features encoding syntax and its linkage to a compositional semantics.[11]

I suggest that "true language" involves the following further properties beyond LR1 through LR7:

LA1. *Symbolization and compositionality*: The symbols become words in the modern sense, interchangeable and composable in the expression of meaning.[12]

LA2. *Syntax, semantics and recursion*: The matching of syntactic to semantic structures coevolves with the fractionation of utterances, with the nesting of substructures making some form of recursion inevitable.

LA1 and LA2 are intertwined. Section 7 will offer candidates for the sorts of discoveries that may have led to progress from "unitary utterances" to more or less structured assemblages of words. Given the view (LR5) that recursion of action (but not of communication) is part of language readiness, the key transition here is the compositionality that allows cognitive structure to be reflected in symbolic structure (the transition from LR2 to LA1), as when perception (not uniquely human) grounds linguistic description (uniquely human) so that, for example, the noun phrase (NP) describing a part of an object may optionally form part of the NP describing the overall object. From this point of view, recursion in language is a corollary of the essentially recursive nature of action and perception *once symbolization becomes compositional, and reflects addition of further detail to, for example, a description when needed to reduce ambiguity in communication*.

The last two principles provide the linguistic complements of two of the conditions for language readiness, LR6 (*Beyond the here-and-now 1*) and LR7 (*Paedomorphy and sociality*), respectively.

LA3. *Beyond the here-and-now 2*: Verb tenses or other circumlocutions express the ability to recall past events or imagine future ones.

There are so many linguistic devices for going beyond the here and now, and beyond the factual, that verb tenses are mentioned to stand in for all the devices languages have developed to communicate about other "possible worlds" that are far removed from the immediacy of, say, the vervet monkey's leopard call.

If one took a human language and removed all reference to time, one might still want to call it a language rather than a protolanguage, even though one would agree that it was thereby greatly impoverished. Similarly, the number system of a language can be seen as a useful, but not definitive, "plug-in." LA3 nonetheless suggests that the ability to talk about past and future is a central part of human languages as we understand them. However, all this would be meaningless (literally) without the underlying cognitive machinery – the substrate for episodic memory provided by the hippocampus (Burgess et al. 1999) and the substrate for planning provided by frontal cortex (Passingham 1993, Ch. 10). It is not part of the mirror system hypothesis to explain the evolution of the brain structures that support LR6; it is an exciting challenge for work "beyond the mirror" to show how such structures could provide the basis for humans to discover the capacities for communication summarized in LA3.

LA4. *Learnability*: To qualify as a human language, much of the syntax and semantics of a human language must be learnable by most human children.

I say "much of" because it is not true that children master all the vocabulary or syntactic subtlety of a language by 5 or 7 years of age. Language acquisition is a process that continues well into the teens as we learn more subtle syntactic expressions and a greater vocabulary to which to apply them (C. Chomsky [1969] traces the changes that occur from ages 5 to 10), allowing us to achieve a richer and richer set of communicative and representational goals.

LR7 and LA4 link a biological condition "orthogonal" to the mirror system hypothesis with a "supplementary" property of human languages. This supplementary property is that languages do not simply exist – they are acquired anew (and may be slightly modified thereby) in each generation (LA4). The biological property is an inherently social one about the nature of the relationship between parent (or other caregiver) and child (LR7) – the prolonged period of infant dependency which is especially pronounced in humans has co-evolved with the social structures for caregiving that provide the conditions for the complex social learning that makes possible the richness of human cultures in general and of human languages in particular (Tomasello 1999b).

2.3. The argument in perspective

The argument unfolds in the remaining six sections as follows:

Section 3. Perspectives on grasping and mirror neurons: This section presents two models of the macaque brain. A key point is that the functions of mirror neurons reflect the impact of experience rather than being pre-wired.

Section 4. Imitation: This section presents the distinction between simple and complex imitation systems for grasping, and argues that monkeys have neither, that chimpanzees have only simple imitation, and that the capacity for complex imitation involved hominid evolution since the separation from our common ancestors, the great apes, including chimpanzees.

Section 5. From imitation to protosign: This section examines the relation between symbolism, intended communication, and parity, and looks at the multiple roles of the mirror system in supporting pantomime and then conventionalized gestures that support a far greater range of intended communication.

Section 6. The emergence of protospeech: This section argues that evolution did not proceed directly from monkey-like primate vocalizations to speech but rather proceeded from vocalization to manual gesture and back to vocalization again.

Section 7. The inventions of languages: This section argues that the transition from action-object frames to verb-argument structures embedded in larger sentences structured by syntax and endowed with a compositional semantics was the effect of the accumulation of a wide range of human discoveries that had little if any impact on the human genome.

Section 8. Toward a neurolinguistics "beyond the mirror": This section extracts a framework for action-oriented linguistics informed by our analysis of the "extended mirror

Table 1. *A comparative view of how the following sections relate the criteria LR1–LR for language readiness and LA1–LA2 for language (middle column) to the seven stages, S1–S7, of the extended mirror system hypothesis (right column)*

Section	Criteria	Stages
2.1	LR5: From hierarchical structuring to temporal ordering	This precedes the evolutionary stages charted here.
3.1		S1: Grasping The FARS model.
3.2		S2: Mirror system for grasping Modeling Development of the Mirror System. This supports the conclusion that mirror neurons can be recruited to recognize and encode an expanding set of novel actions.
4	LR1: Complex imitation	S3: Simple imitation This involves properties of the mirror system beyond the monkey's data. S4: Complex imitation This is argued to distinguish humans from other primates.
5	LR2: Symbolization LR4: Intended communication LR3: Parity (mirror property)	S5: Protosign The transition of complex imitation from praxic to communicative use involves two substages: S5a: the ability to engage in pantomime; S5b: the ability to make conventional gestures to disambiguate pantomime.
6.1		S6: Protospeech It is argued that early protosign provided the scaffolding for early protospeech, after which both developed in an expanding spiral until protospeech became dominant for most people.
7	LA1: Symbolization and compositionality LA2: Syntax, semantics, and recursion	S7: Language The transition from action-object frame to verb-argument structure to syntax and semantics.
8		The evolutionary developments of the preceding sections are restructured into synchronic form to provide a framework for further research in neurolinguistics relating the capabilities of the human brain for language, action recognition, and imitation.

system hypothesis" presented in the previous sections. The language-ready brain contains the evolved mirror system as a key component but also includes many other components that lie outside, though they interact with, the mirror system.

Table 1 shows how these sections relate the evolutionary stages S1 through S7, and their substages, to the above criteria for language readiness and language.[13]

3. Perspectives on grasping and mirror neurons

Mirror neurons in F5, which are active both when the monkey performs certain actions and when the monkey observes them performed by others, are to be distinguished from *canonical neurons* in F5, which are active when the monkey performs certain actions but not when the monkey observes actions performed by others. More subtly, canonical neurons fire when they are presented with a graspable object, irrespective of whether the monkey performs the grasp or not – but clearly this must depend on the extra (inferred) condition that the monkey not only sees the object but is aware, in some sense, that it is possible to grasp it. Were it not for the caveat, canonical neurons would also fire when the monkey observed the object being grasped by another.

The "classic" mirror system hypothesis (sect. 1.2) emphasizes the grasp-related neurons of the monkey premotor area F5 and the homology of this region with human Broca's area. However, Broca's area is part of a larger system supporting language, and so we need to enrich the mirror system hypothesis by seeing how the mirror system for grasping in monkey includes a variety of brain regions in addition to F5. I show this by presenting data and models that locate the canonical system of F5 in a systems perspective (the FARS model of sect. 3.1) and then place the mirror system of F5 in a system perspective (the MNS model of sect. 3.2).

3.1. The FARS model

Given our concern with hand use and language, it is striking that the ability to use the size of an object to preshape the hand while grasping it can be dissociated by brain lesions from the ability to consciously recognize and describe that size. Goodale et al. (1991) studied a patient (D.F.) whose cortical damage allowed signals to flow from primary visual cortex (V1) towards posterior parietal cortex (PP) but not from V1 to inferotemporal cortex (IT). When asked to indicate the width of a single block by means of her index finger and thumb, D.F.'s finger separation bore no relationship to the dimensions of the object and showed considerable trial-to-trial variability. Yet when she was asked simply to reach out and pick up the block, the peak aperture (well before contact with the object) between her index finger and thumb changed systematically with the width of the object, as in normal controls. A similar dissociation was seen in her responses to the orientation of stimuli. In other words, D.F. could preshape accurately, even though she appeared to have no conscious appreciation (expressible either verbally or in pantomime) of the visual parameters that guided the preshape. Jeannerod et al. (1994) reported a study of impairment of grasping in a patient (A.T.) with a bilateral posterior parietal lesion of vascular origin that left IT and the pathway V1 → IT relatively intact, but grossly impaired the pathway V1 → PP. This patient can reach without deficit toward the location of such an object, but cannot preshape appropriately when asked to grasp it.

A corresponding distinction in the role of these pathways in the monkey is crucial to the FARS model (named for Fagg, Arbib, Rizzolatti, and Sakata; see Fagg & Arbib 1998), which embeds F5 canonical neurons in a larger system. Taira et al. (1990) found that anterior intraparietal (AIP) cells (in the anterior intraparietal sulcus of the parietal cortex) extract neural codes for *affordances* for grasping from the visual stream and sends these on to area F5. Affordances (Gibson 1979) are features of the object relevant to action, in this case to grasping, rather than aspects of identifying the object's identity. Turning to human data: Ehrsson et al. (2003) compared the brain activity when humans attempted to lift an immovable test object held between the tips of the right index finger and thumb with the brain activity obtained in two control tasks in which neither the load force task nor the grip force task involved coordinated grip-load forces. They found that the grip-load force task was specifically associated with activation of a section of the right intraparietal cortex. Culham et al. (2003) found greater activity for grasping than for reaching in several regions, including the anterior intraparietal (AIP) cortex. Although the lateral occipital complex (LOC), a ventral stream area believed to play a critical role in object recognition, was activated by the objects presented on both grasping and reaching trials, there was no greater activity for grasping compared to reaching.

The FARS model analyzes how the "canonical system," centered on the AIP → F5 pathway, may account for basic phenomena of grasping. The highlights of the model are shown in Figure 2,[14] which diagrams the crucial role of IT (inferotemporal cortex) and PFC (prefrontal cortex) in modulating F5's selection of an affordance. The *dorsal stream* (from V1 to parietal cortex) carries the information needed for AIP to recognize that different parts of the object can be grasped in different ways, thus extracting affordances for the grasp system which are then passed on to F5. The dorsal stream does not know "what" the object is; it can only see the object as a set of possible affordances. The *ventral stream* (from V1 to IT), by contrast, is able to recognize what the object is. This information is passed to PFC, which can then, on the basis of the current goals of the organism and the recognition of the nature of the object, bias AIP to choose the affordance appropriate to the task at hand. The original FARS model posited connections between PFC and F5. However, there is evidence (reviewed by Rizzolatti & Luppino 2001) that these connections are very limited, whereas rich connections exist between PFC and AIP. Rizzolatti and Luppino (2003) therefore suggested that FARS

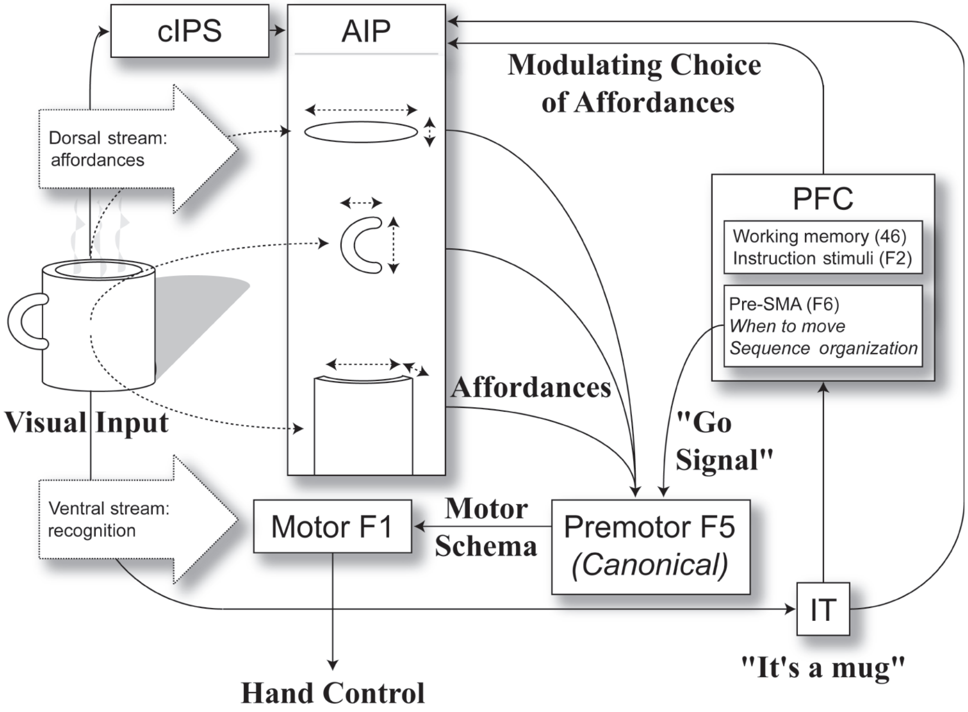


Figure 2. A reconceptualization of the FARS model in which the primary influence of PFC (prefrontal cortex) on the selection of affordances is on parietal cortex (AIP, anterior intraparietal sulcus) rather than premotor cortex (the hand area F5).

be modified so that information on object semantics and the goals of the individual influence AIP rather than F5 neurons. I show the modified schematic in Figure 2. The modified figure represents the way in which AIP may accept signals from areas F6 (pre-SMA), 46 (dorsolateral prefrontal cortex), and F2 (dorsal premotor cortex) to respond to task constraints, working memory, and instruction stimuli, respectively. In other words, AIP provides cues on how to interact with an object, leaving it to IT to categorize the object or determine its identity.

Although the data on cell specificity in F5 and AIP emphasize single actions, these actions are normally part of more complex behaviors – to take a simple example, a monkey who grasps a raisin will, in general, then proceed to eat it. Moreover, a particular action might be part of many learned sequences, and so we do not expect the premotor neurons for one action to prime a single possible consequent action and hence must reject "hard wiring" of the sequence. The generally adopted solution is to segregate the learning of a sequence from the circuitry which encodes the unit actions, the latter being F5 in the present study. Instead, another area (possibly the part of the supplementary motor area called pre-SMA; Rizzolatti et al. 1998) has neurons whose connections encode an "abstract sequence" Q1, Q2, Q3, Q4, with sequence learning then involving learning that the activation of Q1 triggers the F5 neurons for A, Q2 triggers B, Q3 triggers A again, and Q4 triggers C to provide encoding of the sequence A-B-A-C. Other studies suggest that administration of the sequence (inhibiting extraneous actions, while priming imminent actions) is carried out by the basal ganglia on the basis of its interactions with the pre-SMA (Bischoff-Grethe et al. 2003; see Dominey et al. 1995 for an earlier model of the possible role of the basal ganglia in sequence learning).

3.2. Modeling development of the mirror system

The populations of canonical and mirror neurons appear to be spatially segregated in F5 (Rizzolatti & Luppino 2001). Both sectors receive a strong input from the secondary somatosensory area (SII) and parietal area PF. In addition, canonical neurons are the selective target of area AIP. Perrett et al. (1990; cf. Carey et al. 1997) found that STSa, in the rostral part of the superior temporal sulcus (STS), has neurons which discharge when the monkey observes such biological actions as walking, turning the head, bending the torso, and moving the arms. Of most relevance to us is that a few of these neurons discharged when the monkey observed goal-directed hand movements, such as grasping objects (Perrett et al. 1990) – though STSa neurons do not seem to discharge during movement execution as distinct from observation. STSa and F5 may be indirectly connected via the inferior parietal area PF (Brodmann area 7b) (Cavada & Goldman-Rakic 1989; Matelli et al. 1986; Petrides & Pandya 1984; Seltzer & Pandya 1994). About 40% of the visually responsive neurons in PF are active for observation of actions such as holding, placing, reaching, grasping, and bimanual interaction. Moreover, most of these action-observation neurons were also active during the execution of actions similar to those for which they were "observers," and were therefore called PF mirror neurons (Fogassi et al. 1998).

In summary, area F5 and area PF include an observation/execution matching system: When the monkey observes an action that resembles one in its movement repertoire, a subset of the F5 and PF mirror neurons is activated which also discharges when a similar action is executed by the monkey itself.

I next develop the conceptual framework for thinking about the relation between F5, AIP, and PF. Section 6.1 expands the mirror neuron database, reviewing the reports by Kohler et al. (2002) of a subset of mirror neurons responsive to sounds and by Ferrari et al. (2003) of neurons responsive to the observation of orofacial communicative gestures.

Figure 3 provides a glimpse of the schemas (functions) involved in the MNS model (Oztop & Arbib 2002) of the monkey mirror system.[15] First, we look at those elements involved when the monkey itself reaches for an object. Areas IT and cIPS (caudal intraparietal sulcus; part of area 7) provide visual input concerning the nature of the observed object and the position and orientation of the object's surfaces, respectively, to AIP. The job of AIP is then to extract the affordances the object offers for grasping. The upper diagonal in Figure 3 corresponds to the basic pathway AIP → $F5_{canonical}$ → M1 (primary motor cortex) of the FARS model, but Figure 3 does not include the important role of PFC in action selection. The lower-right diagonal (MIP/LIP/VIP → F4) completes the "canonical" portion of the MNS model, since motor cortex must instruct not only the hand muscles how to grasp but also (via various intermediaries) the arm muscles how to reach, transporting the hand to the object. The rest of Figure 3 presents the core elements for the understanding of the mirror system. Mirror neurons do not fire when the monkey sees the hand movement or the object in isolation – it is the sight of the hand moving appropriately to grasp or otherwise manipulate a seen (or recently seen) object (Umiltá et al. 2001) that is required for the mirror neurons attuned to the given action to fire. This requires schemas for the recognition of both the shape of the hand and analysis of its motion (ascribed in the figure to STS), and for analysis of the relation of these hand parameters to the location and affordance of the object (7a and 7b; we identify 7b with PF).

In the MNS model, the *hand state* was accordingly defined as a vector whose components represented the movement of the wrist relative to the location of the object and of the hand shape relative to the affordances of the object. Oztop and Arbib (2002) showed that an artificial neural network corresponding to PF and $F5_{mirror}$ could be trained to recognize the grasp type from the *hand state trajectory*, with correct classification often being achieved well before the hand reached the object. The modeling assumed that the neural equivalent of a grasp being in the monkey's repertoire is that there is a pattern of activity in the F5 canonical neurons which commands that grasp. During training, the output of the F5 canonical neurons, acting as a code for the grasp being executed by the monkey at that time, was used as the training signal for the F5 mirror neurons to enable them to learn which hand-object trajectories corresponded to the canonically encoded grasps. Moreover, the input to the F5 mirror neurons encodes the trajectory of the relation of parts of the hand to the object rather than the visual appearance of the hand in the visual field. As a result of this training, the appropriate mirror neurons come to fire in response to viewing the appropriate trajectories even when the trajectory is not accompanied by F5 canonical firing.

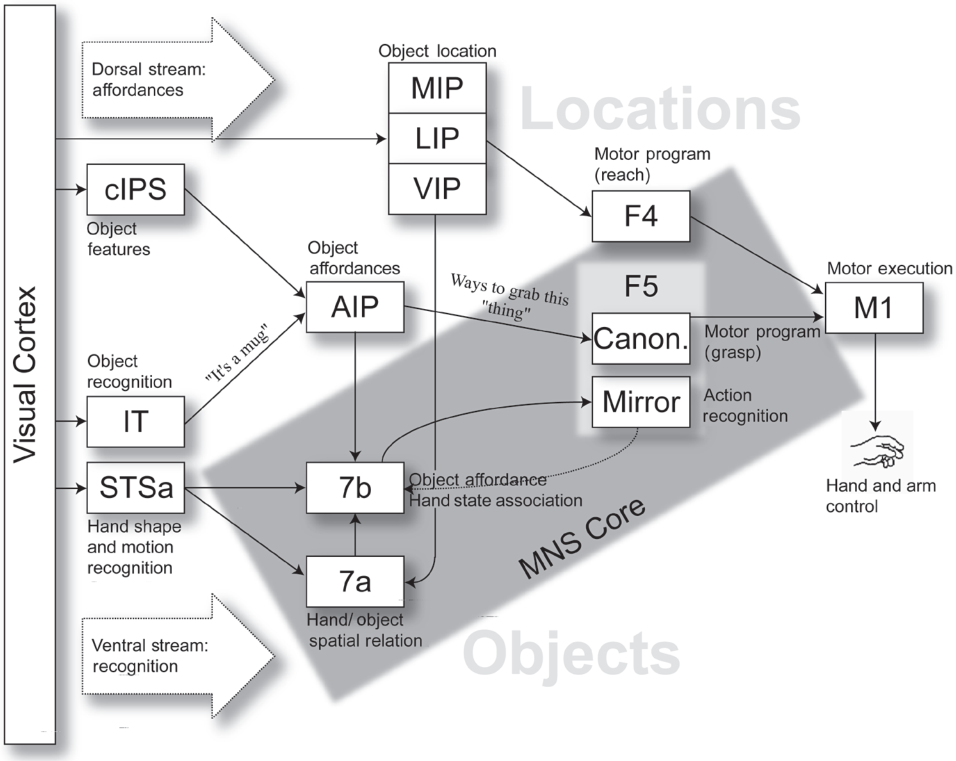


Figure 3. A schematic view of the Mirror Neuron System (MNS) model (Oztop & Arbib 2002).

This training prepares the F5 mirror neurons to respond to hand-object relational trajectories even when the hand is of the "other" rather than the "self," because the hand state is based on the movement of a hand relative to the object, and thus only *indirectly* on the retinal input of seeing hand and object which can differ greatly between observation of self and other. What makes the modeling worthwhile is that the trained network not only responded to hand-state trajectories from the training set, but also exhibited interesting responses to novel hand-object relationships. Despite the use of a non-physiological neural network, simulations with the model revealed a range of putative properties of mirror neurons that suggest new neurophysiological experiments. (See Oztop & Arbib [2002] for examples and detailed analysis.)

Although MNS was constructed as a model of the development of mirror neurons in the monkey, it serves equally well as a model of the development of mirror neurons in the human infant. A major theme for future modeling, then, will be to clarify which aspects of human development are generic for primates and which are specific to the human repertoire. In any case, the MNS model makes the crucial assumption that the grasps that the mirror system comes to recognize are already in the (monkey or human) infant's repertoire. But this raises the question of how grasps entered the repertoire. To simplify somewhat, the answer has two parts: (1) Children explore their environment, and as their initially inept arm and hand movements successfully contact objects, they learn to reproduce the successful grasps reliably, with the repertoire being tuned through further experience. (2) With more or less help from caregivers, infants come to recognize certain novel actions in terms of similarities with and differences from movements already in their repertoires, and on this basis learn to produce some version of these novel actions for themselves. Our Infant Learning to Grasp Model (ILGM; Oztop et al. 2004) strongly supports the hypothesis that grasps are acquired through experience as the infant learns how to conform the biomechanics of its hand to the shapes of the objects it encounters. However, limited space precludes presentation of this model here.

The classic papers on the mirror system for grasping in the monkey focus on a repertoire of grasps – such as the precision pinch and power grasp – that seem so basic that it is tempting to think of them as prewired. The crucial point of this section on modeling is that learning models such as ILGM and MNS, and the data they address, make clear that *mirror neurons are not restricted to recognition of an innate set of actions but can be recruited to recognize and encode an expanding repertoire of novel actions*. I will relate the FARS and MNS models to the development of imitation at the end of section 4.

With this, let us turn to human data. We mentioned in section 1.2 that Broca's area, traditionally thought of as a speech area, has been shown by brain imaging studies to be active when humans both execute and observe grasps. This was first tested by two positron emission tomography (PET) experiments (Grafton et al. 1996; Rizzolatti et al. 1996) which compared brain activation when subjects observed the experimenter grasping an object against activation when subjects simply observed the object. Grasp observation significantly activated the superior temporal sulcus (STS), the inferior parietal lobule, and the inferior frontal gyrus (area 45). All activations were in the left hemisphere. The last area is of especial interest because areas 44 and 45 in the left hemisphere of the human constitute Broca's area. Such data certainly contribute to the growing body of indirect evidence that there is a mirror system for grasping that links Broca's area with regions in the inferior parietal lobule and STS. We have seen that the "minimal mirror system" for grasping in the macaque includes mirror neurons in the parietal area PF (7b) as well as F5, and some not-quite-mirror neurons in the region STSa in the superior temporal sulcus. Hence, in further investigation of the mirror system hypothesis it will be crucial to extend the F5 → Broca's area homology to examine the human homologues of PF and STSa as well. I will return to this issue in section 7 (see Fig. 6) and briefly review some of the relevant data from the rich and rapidly growing literature based on human brain imaging and transcranial magnetic stimulation (TMS) inspired by the effort to probe the human mirror system and relate it to action recognition, imitation, and language.

Returning to the term "language readiness," let me stress that the reliable linkage of brain areas to different aspects of language in normal speaking humans does not imply that language per se is "genetically encoded" in these regions. There is a neurology of writing even though writing was invented only a few thousand years ago. The claim is not that Broca's area, Wernicke's area, and STS are genetically preprogrammed for language, but rather that the development of a human child in a language community normally adapts these brain regions to play a crucial (but not the only) role in language performance.

4. Imitation

We have already discussed the mirror system for grasping as something shared between macaque and human; hence the hypothesis that this set of mechanisms was already in place in the common ancestor of monkey and human some 20 million years ago.[16] In this section we move from stage S2, a mirror system for grasping, to stages S3, a simple imitation system for grasping, and S4, a complex imitation system for grasping. I will argue that chimpanzees possess a capability for *simple* imitation that monkeys lack, but that humans have *complex* imitation whereas other primates do not. The ability to copy *single* actions is just the first step towards complex imitation, which involves parsing a complex movement into more or less familiar pieces and then performing the corresponding composite of (variations on) familiar actions. Arbib and Rizzolatti (1997) asserted that what makes a movement into an action is that it is associated with a goal, and that initiation of the movement is accompanied by the creation of an expectation that the goal will be met. Hence, it is worth stressing that when I speak of imitation here, I speak of the imitation of a movement and its linkage to the goals it is meant to achieve. The action may thus vary from occasion to occasion depending on parametric variations in the goal. This is demonstrated by Byrne's (2003) description, noted earlier, of a mountain gorilla preparing bundles of nettle leaves to eat.

Visalberghi and Fragaszy (2002) review data on attempts to observe imitation in monkeys, including their own studies of capuchin monkeys. They stress the huge difference between the major role that imitation plays in learning by human children, and the very limited role, if any, that imitation plays in social learning in monkeys. There is little evidence for vocal imitation in monkeys or apes (Hauser 1996), but it is generally accepted that chimpanzees are capable of some forms of imitation (Tomasello & Call 1997).

There is not space here to analyze all the relevant distinctions between imitation and other forms of learning, but one example may clarify my view: Voelkl and Huber (2000) had marmosets observe a demonstrator removing the lids from a series of plastic canisters to obtain a mealworm. When subsequently allowed access to the canisters, marmosets that observed a demonstrator using its hands to remove the lids used only their hands. In contrast, marmosets that observed a demonstrator using its mouth also used their mouths to remove the lids. Voelkl and Huber (2000) suggest that this may be a case of true imitation in marmosets, but I would argue that it is a case of *stimulus enhancement*, apparent imitation resulting from directing attention to a particular object or part of the body or environment. This is to be distinguished from *emulation* (observing and attempting to reproduce results of another's actions without paying attention to details of the other's behavior) and *true imitation* which involves copying a novel, otherwise improbable action or some act that is outside the imitator's prior repertoire.

Myowa-Yamakoshi and Matsuzawa (1999) observed in a laboratory setting that chimpanzees typically took 12 trials to learn to "imitate" a behavior and in doing so paid more attention to where the manipulated object was being directed than to the actual movements of the demonstrator. This involves the ability to learn novel actions which may require using one or both hands to bring two objects into relationship, or to bring an object into relationship with the body.

Chimpanzees do use and make tools in the wild, with different tool traditions found in geographically separated groups of chimpanzees: Boesch and Boesch (1983) have observed chimpanzees in Tai National Park, Ivory Coast, using stone tools to crack nuts open, although Goodall has never seen chimpanzees do this in the Gombe in Tanzania. They crack harder-shelled nuts with stone hammers and stone anvils. The Tai chimpanzees live in a dense forest where suitable stones are hard to find. The stone anvils are stored in particular locations to which the chimpanzees continually return.[17] The nut-cracking technique is not mastered until adulthood. Tomasello (1999b) comments that, over many years of observation, Boesch observed only two possible instances in which the mother *appeared* to be actively attempting to instruct her child, and that even in these cases it is unclear whether the mother had the goal of helping the young chimp learn to use the tool. We may contrast the long and laborious process of acquiring the nut-cracking technique with the rapidity with which human adults can acquire novel sequences, and the crucial role of caregivers in the development of this capacity for complex imitation. Meanwhile, reports abound of imitation in many species, including dolphins and orangutans, and even tool use in crows (Hunt & Gray 2002). Consequently, I accept that the demarcation between the capability for imitation of humans and nonhumans is problematic. Nonetheless, I still think it is fair to claim that humans can master feats of imitation beyond those possible for other primates.

The ability to imitate has clear adaptive advantage in allowing creatures to transfer skills to their offspring, and therefore could be selected for quite independently of any adaptation related to the later emergence of protolanguage. By the same token, the ability for complex imitation could provide further selective advantage unrelated to language. However, complex imitation is central to human infants both in their increasing mastery of the physical and social world and in the close coupling of this mastery to the acquisition of language (cf. Donald 1998; Arbib et al., in press). The child must go beyond simple imitation to acquire the phonological repertoire, words, and basic "assembly skills" of its language community, and this is one of the ways in which brain mechanisms supporting imitation were crucial to the emergence of language-ready *Homo sapiens*. If I then assume (1) that the common ancestor of monkeys and apes had no greater imitative ability than present-day monkeys (who possess, I suggest, stimulus enhancement rather than simple imitation), and (2) that the ability for simple imitation shared by chimps and humans was also possessed by their common ancestor, but (3) that only humans possess a talent for "complex" imitation, then I have established a case for the hypothesis that extension of the mirror system from *recognizing* single actions to *being able to copy* compound actions was the key innovation in the brains of our hominid ancestors that was relevant to language. And, more specifically, we have the hypotheses:

Stage S3 hypothesis: Brain mechanisms supporting a simple imitation system – imitation of short, novel sequences of object-directed actions through repeated exposure – for grasping developed in the 15-million-year evolution from the common ancestor of monkeys and apes to the common ancestor of apes and humans; and

Stage S4 hypothesis: Brain mechanisms supporting a complex imitation system – acquiring (longer) novel sequences of more abstract actions in a single trial – developed in the 5-million-year evolution from the common ancestor of apes and humans along the hominid line that led, in particular, to *Homo sapiens*.[18]

Now that we have introduced imitation, we can put the models of section 3.2 in perspective by postulating the following stages prior to, during, and building on the development of the mirror system for grasping in the infant:

A. The child refines a crude map (superior colliculus) to make unstructured reach and "swipe" movements at objects; the grasp reflex occasionally yields a successful grasp.

B. The child develops a set of grasps which succeed by kinesthetic, somatosensory criteria (ILGM).

C. AIP develops as affordances of objects become learned in association with successful grasps. Grasping becomes visually guided; the grasp reflex disappears.

D. The (grasp) mirror neuron system develops driven by visual stimuli relating hand and object generated by the actions (grasps) performed by the infant himself (MNS).

E. The child gains the ability to map other individual's actions into his internal motor representation.

F. Then the child acquires the ability to imitate, creating (internal) representations for novel actions that have been observed and developing an action prediction capability.

I suggest that stages A through D are much the same in monkey and human, but that stages E and F are rudimentary at best in monkeys, somewhat developed in chimps, and well-developed in human children (but not in infants). In terms of Figure 3, we might say that if MNS were augmented to have a population of mirror neurons that could acquire population codes for observed actions not yet in the repertoire of self-actions, then in stage E the mirror neurons would provide training for the canonical neurons, reversing the information flow seen in the MNS model. Note that this raises the further possibility that the human infant may come to recognize movements that not only are not within the repertoire but which never come to be within the repertoire. In this case, the cumulative development of action recognition may proceed to increase the breadth and subtlety of the range of actions that are recognizable but cannot be performed by children.

5. From imitation to protosign

The next posited transition, from stage S4, a complex imitation system for grasping, to stage S5, protosign, a manual-based communication system, takes us from imitation for the sake of instrumental goals to imitation for the sake of communication. Each stage builds on, yet is not simply reducible to, the previous stage.

I argue that the combination of the abilities (S5a) to engage in pantomime and (S5b) to make conventional gestures to disambiguate pantomime yielded a brain which could (S5) support "protosign," a manual-based communication system that broke through the fixed repertoire of primate vocalizations to yield an open repertoire of communicative gestures.

It is important to stress that communication is about far more than grasping. To pantomime the flight of a bird, you might move your hand up and down in a way that indicates the flapping of a wing. Your pantomime uses movements of the hand (and arm and body) to imitate movement other than hand movements. You can pantomime an object either by miming a typical action by or with the object, or by tracing out the characteristic shape of the object.

The transition to pantomime does seem to involve a genuine neurological change. Mirror neurons for grasping in the monkey will fire only if the monkey sees *both* the hand movement and the object to which it is directed (Umiltá et al. 2001). A grasping movement that is not made in the presence of a suitable object, or is not directed toward that object, will not elicit mirror neuron firing. By contrast, in pantomime, the observer sees the movement in isolation and *infers* (1) what non-hand movement is being mimicked by the hand movement, and (2) the goal or object of the action. This is an evolutionary change of key relevance to language readiness. Imitation is the generic attempt to reproduce movements performed by another, whether to master a skill or simply as part of a social interaction. By contrast, pantomime is performed with the intention of getting the observer to think of a specific action, object, or event. It is essentially communicative in its nature. The imitator observes; the pantomimic intends to be observed.

As Stokoe (2001) and others emphasize, the power of pantomime is that it provides open-ended communication that works without prior instruction or convention. However (and I shall return to this issue at the end of this section), even signs of modern signed language which resemble pantomimes are conventionalized and are, thus, distinct from pantomimes. Pantomime per se is not a form of protolanguage; rather it provides a rich scaffolding for the emergence of protosign.

All this assumes rather than provides an explanation for LR4, the transition from making praxic movement – for example, those involved in the immediate satisfaction of some appetitive or aversive goal – to those intended by the utterer to have a particular effect on the recipient. I tentatively offer:

The intended communication hypothesis: The ability to imitate combines with the ability to observe the effect of such imitation on conspecifics to support a migration of closed species-specific gestures supported by other brain regions to become the core of an open class of communicative gestures.

Darwin (1872/1965) observed long ago, across a far wider range of mammalian species than just the primates, that the facial expressions of conspecifics provide valuable cues to their likely reaction to certain courses of behavior (a rich complex summarized as "emotional state"). Moreover, the F5 region contains orofacial cells as well as manual cells. This suggests a progression from control of emotional expression by systems that *exclude* F5 to the extension of F5's mirror capacity for orofacial as well as manual movement (discussed below), via its posited capacity (achieved by stage S3) for simple imitation, to support the imitation of emotional expressions. This would then provide the ability to affect the behavior of others by, for example, *appearing* angry. This would in turn provide the evolutionary opportunity to generalize the ability of F5 activity to affect the behavior of conspecifics from species-specific vocalizations to a general ability to use the imitation of behavior (as distinct from praxic behavior itself) as a means to influence others. This in turn makes possible reciprocity by a process of backward chaining where the influence is not so much on the praxis of the other as on the exchange of information. With this, the transition described by LR4 (intended communication) has been achieved in tandem with the achievement and increasing sophistication of LR2 (symbolization).

A further critical change (labeled 5b above) emerges from the fact that in pantomime it might be hard to distinguish, for example, a movement signifying "bird" from one meaning "flying." This inability to adequately convey shades of meaning using "natural" pantomime would favor the invention of gestures that could in some way disambiguate which of their associated meanings was intended. Note that whereas a pantomime can freely use any movement that might evoke the intended observation in the mind of the observer, a disambiguating gesture must be conventionalized.[19] This use of non-pantomimic gestures requires extending the use of the mirror system to attend to an entirely new class of hand movements. However, this does not seem to require a biological change beyond that limned above for pantomime.

As pantomime begins to use hand movements to mime different degrees of freedom (as in miming the flying of a bird), a dissociation begins to emerge. The mirror system for the pantomime (based on movements of face, hand, etc.) is now different from the recognition system for the action that is pantomimed, and – as in the case of flying – the action may not even be in the human action repertoire. However, the system is still able to exploit the praxic recognition system because an animal or hominid must observe much about the environment that is relevant to its actions but is not in its own action repertoire. Nonetheless, this dissociation now underwrites the emergence of protosign – an open system of actions that are defined only by their communicative impact, not by their direct relation to praxic goals.

Protosign may lose the ability of the original pantomime to elicit a response from someone who has not seen it before. However, the price is worth paying in that the simplified form, once agreed upon by the community, allows more rapid communication with less neural effort. One may see analogies in the history of Chinese characters. The character 山 (san) may not seem particularly pictorial, but if (following the "etymology" of Vaccari & Vaccari 1961), we see it as a simplification of a picture of three mountains, , via such intermediate forms as , then we have no trouble seeing the simplified character 山 as meaning "mountain."[20] The important point here for our hypothesis is that although such a "picture history" may provide a valuable crutch to some learners, with sufficient practice the crutch is thrown away, and in normal reading and writing, the link between 山 and its meaning is direct, with no need to invoke an intermediate representation of .

In the same way, I suggest that pantomime is a valuable crutch for acquiring a modern sign language, but that even signs which resemble pantomimes are conventionalized and are thus distinct from pantomimes.[21] Interestingly, Emmorey (2002, Ch. 9) discusses studies of signers using ASL which show a dissociation between the neural systems involved in sign language and those involved in conventionalized gesture and pantomime. Corina et al. (1992b) reported left-hemisphere dominance for producing ASL signs, but no laterality effect when subjects had to produce symbolic gestures (e.g., waving good-bye or thumbs-up). Other studies report patients with left-hemisphere damage who exhibited sign language impairments but well-preserved conventional gesture and pantomime. Corina et al. (1992a) described patient W.L. with damage to left-hemisphere perisylvian regions. W.L. exhibited poor sign language comprehension and production. Nonetheless, this patient could produce stretches of pantomime and tended to substitute pantomimes for signs, even when the pantomime required more complex movement. Emmorey sees such data as providing neurological evidence that signed languages consist of linguistic gestures and not simply elaborate pantomimes.

Figure 4 is based on a scheme offered by Arbib (2004) in response to Hurford's (2004) critique of the mirror system hypothesis. Hurford makes the crucial point that we must (in the spirit of Saussure) distinguish the "sign" from the "signified." In the figure, we distinguish the "neural representation of the sign" (top row) from the "neural representation of the signified" (bottom row). The top row of the figure makes explicit the result of the progression within the mirror system hypothesis of mirror systems for:

1. Grasping and manual praxic actions.
2. Pantomime of grasping and manual praxic actions.
3. Pantomime of actions outside the pantomimic's own behavioral repertoire (e.g., flapping the arms to mime a flying bird).
4. Conventional gestures used to formalize and disambiguate pantomime (e.g., to distinguish "bird" from "flying").
5. Protosign, comprising conventionalized manual (and related orofacial) communicative gestures.

However, I disagree with Hurford's suggestion that there is a mirror system for all concepts – actions, objects, and more – which links the perception and action related to each concept.[22] In schema theory (Arbib 1981; 2003), I distinguish between *perceptual schemas*, which determine whether a given "domain of interaction" is present in the environment and provide parameters concerning the current relationship of the organism with that domain, and *motor schemas*, which provide the control systems which can be coordinated to effect a wide variety of actions. Recog-

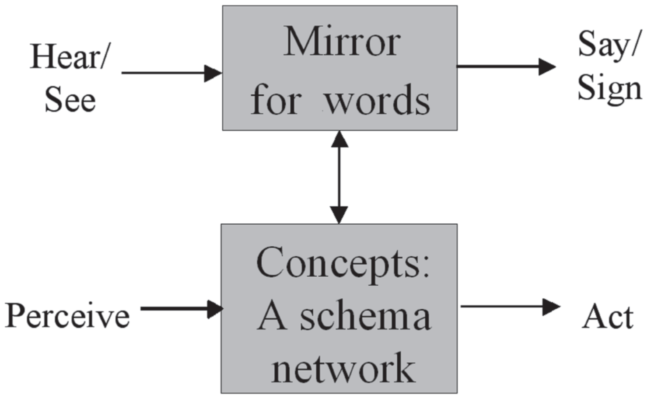


Figure 4. The bidirectional sign relation links words and concepts. The top row concerns Phonological Form, which may relate to signed language as much as to spoken language. The bottom row concerns Cognitive Form and includes the recognition of objects and actions. Phonological Form is present only in humans while Cognitive Form is present in both monkeys and humans. The Mirror System Hypothesis hypothesizes that there is a mirror system for words, but there may not be a mirror system for concepts.

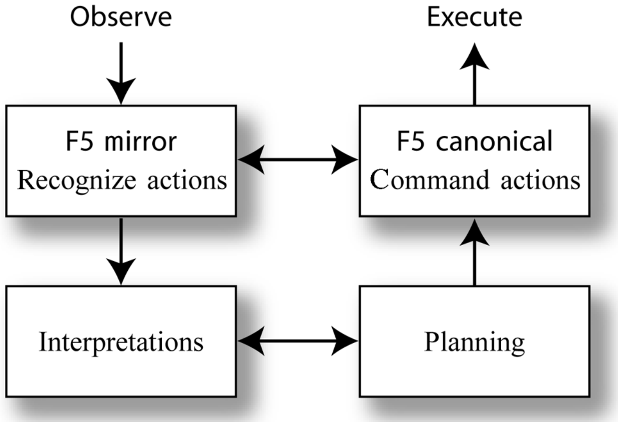


Figure 5. The perceptuomotor coding for both observation and execution contained in the mirror system for manual actions in the monkey is linked to "conceptual systems" for interpretation and planning of such actions. The interpretation and planning systems themselves do not have the mirror property save through their linkage to the actual mirror system.

nizing an object (an apple, say) may be linked to many different courses of action (to place the apple in one's shopping basket; to place the apple in the bowl at home; to peel the apple; to eat the apple; to discard a rotten apple, etc.). In this list, some items are apple-specific, whereas other invoke generic schemas for reaching and grasping. Such considerations led me to separate perceptual and motor schemas – a given action may be invoked in a wide variety of circumstances; a given perception may, as part of a larger assemblage, precede many courses of action. Hence, I reject the notion of a mirror system for concepts. Only rarely (as in the case of certain basic actions such as *grasp* or *run*, or certain expressions of emotion) will the perceptual and motor schemas be integrated into a "mirror schema." I do not see a "concept" as corresponding to one word, but rather to a graded set of activations of the schema network.

But if this is the case, does a mirror system for protosigns (and, later, for the words and utterances of a language) really yield the LR3 form of the mirror property – that what counts for the sender must count for the receiver? Actually, it yields only half of this directly: the recognition that the action of the observed protosigner is his or her version of one of the conventional gestures in the observer's repertoire. The claim, then, is that the LR3 form of the mirror property – that which counts for the sender must count for the receiver – does not result from the evolution of the F5 mirror system *in and of itself* to support communicative gestures as well as praxic actions; rather, this evolution occurs within the neural context that links the execution and observation of an action to the creature's planning of its own actions and interpretations of the actions of others (Fig. 5). These linkages extract more or less coherent patterns from the creature's experience of the effects of its own actions as well as the consequences of actions by others. Similarly, execution and observation of a communicative action must be linked to the creature's planning and interpretations of communication with others in relation to the ongoing behaviors that provide the significance of the communicative gestures involved.

6. The emergence of protospeech

6.1. The path to protospeech is indirect

My claim here is that the path to protospeech is indirect, with early protosign providing a necessary scaffolding for the emergence of protospeech. I thus reject the claim that speech evolved directly as an elaboration of a closed repertoire of alarm calls and other species-specific vocalizations such as exhibited by nonhuman primates. However, I claim neither that protosign attained the status of a full language prior to the emergence of early forms of protospeech, nor even that stage S5 (protosign) was completed before stage S6 (protospeech) began.

Manual gesture certainly appears to be more conducive to iconic representation than oral gesture. The main argument of section 5 was that the use of pantomime made it easy to acquire a core vocabulary, while the discovery of a growing stock of conventional signs (or sign modifiers) to mark important distinctions then created a culture in which the use of arbitrary gestures would increasingly augment and ritualize (without entirely supplanting) the use of pantomime.[23] Once an organism has an iconic gesture, it can both modulate that gesture and/or symbolize it (non-iconically) by "simply" associating a vocalization with it. Once the association had been learned, the "scaffolding" gesture (like the pantomime that supported its conventionalization, or the caricature that supports the initial understanding of some Chinese ideograms) could be dropped to leave a symbol that need have no remaining iconic relation to its referent, even if the indirect associative relationship can be recalled on some occasions. One open question is the extent to which protosign must be in place before this scaffolding can effectively support the development of protospeech. Because there is no direct mapping of sign (with its use of concurrency and signing space) to phoneme sequences, I think that this development is far more of a breakthrough than it may at first sight appear.

I have separated S6, the evolution of protospeech, from S5, the evolution of protosign, to stress the point that the role of F5 in grounding the evolution of a protolanguage system would work just as well if we and all our ancestors had been deaf. However, primates do have a rich auditory system which contributes to species survival in many ways, of which communication is just one (Ghazanfar 2003). The protolanguage perception system could thus build upon the existing auditory mechanisms in the move to derive protospeech. However, it appears that considerable evolution of the vocal-motor system was needed to yield the flexible vocal apparatus that distinguishes humans from other primates. MacNeilage (1998) offers an argument for how the mechanism for producing consonant-vowel alternations en route to a flexible repertoire of syllables might have evolved from the cyclic mandibular alternations of eating, but offers no clue as to what might have linked such a process to the expression of meaning (but see MacNeilage & Davis, in press b). This problem is discussed much further in Arbib (2005) which spells out how protosign (S5) may have provided a scaffolding for protospeech (S6), forming an "expanding spiral" wherein the two interacted with each other in supporting the evolution of brain and body that made *Homo sapiens* "language-ready" in a multi-modal integration of manual, facial and vocal actions.

New data on mirror neurons for grasping that exhibit auditory responses, and on mirror-like properties of orofacial neurons in F5, add to the subtlety of the argument. Kohler et al. (2002) studied mirror neurons for actions which are accompanied by characteristic sounds, and found that a subset of these neurons are activated by the sound of the action (e.g., breaking a peanut in half) as well as sight of the action. Does this suggest that protospeech mediated by the F5 homologue in the hominid brain could have evolved without the scaffolding provided by protosign? My answer is negative for two reasons: (1) I have argued that imitation is crucial to grounding pantomime in which a movement is performed in the absence of the object for which such a movement would constitute part of a praxic action. However, the sounds studied by Kohler et al. (2002) cannot be created in the absence of the object, and there is no evidence that monkeys can use their vocal apparatus to mimic the sounds they have heard. I would further argue that the limited number and congruence of these "auditory mirror neurons" is more consistent with the view that manual gesture is primary in the early stages of the evolution of language readiness, with audiomotor neurons laying the basis for later extension of protosign to protospeech.

Complementing earlier studies on hand neurons in macaque F5, Ferrari et al. (2003) studied mouth motor neurons in F5 and showed that about one-third of them also discharge when the monkey observes another individual performing mouth actions. The majority of these "mouth mirror neurons" become active during the execution and observation of mouth actions related to ingestive functions such as grasping, sucking, or breaking food. Another population of mouth mirror neurons also discharges during the execution of ingestive actions, but the most effective visual stimuli in triggering them are communicative mouth gestures (e.g., lip-smacking) – one action becomes associated with a whole performance of which one part involves similar movements. This fits with the hypothesis that neurons learn to associate patterns of neural firing rather than being committed to learn specifically pigeonholed categories of data. Thus, a potential mirror neuron is in no way committed to become a mirror neuron in the strict sense, even though it may be more likely to do so than otherwise. The observed communicative actions (with the effective executed action for different "mirror neurons" in parentheses) include lip-smacking (sucking and lip-smacking); lips protrusion (grasping with lips, lips protrusion, lip-smacking, grasping, and chewing); tongue protrusion (reaching with tongue); teeth-chatter (grasping); and lips/tongue protrusion (grasping with lips and reaching with tongue; grasping). We therefore see that the communicative gestures and their associated effective observed actions are a long way from the sort of vocalizations that occur in speech (see Fogassi & Ferrari [in press] for further discussion).

Rizzolatti and Arbib (1998) stated that "This new use of vocalization [in speech] necessitated its skillful control, a requirement that could not be fulfilled by the ancient emotional vocalization centers. This new situation was most likely the 'cause' of the emergence of human Broca's area." I would now rather say that *Homo habilis* and even more so *Homo erectus* had a "proto-Broca's area" based on an F5-like precursor mediating communication by manual and orofacial gestures, which made possible a process of collateralization whereby this "proto" Broca's area gained primitive control of the vocal machinery, thus yielding increased skill and openness in vocalization, moving from the fixed repertoire of primate vocalizations to the unlimited (open) range of vocalizations exploited in speech. Speech apparatus and brain regions could then coevolve to yield the configuration seen in modern *Homo sapiens*.

Corballis (2003b) argues that there may have been a single-gene mutation producing a "dextral" allele, which created a strong bias toward right-handedness and left-cerebral dominance for language at some point in hominid evolution.[24] He then suggests that the "speciation event" that distinguished *Homo sapiens* from other large-brained hominids may have been a switch from a predominantly gestural to a predominantly vocal form of language. By contrast, I would argue that there was no one distinctive speciation event, and that the process whereby communication for most humans became predominantly vocal was not a switch but was "cultural" and cumulative.

7. The inventions of languages

The divergence of the Romance languages from Latin took about one thousand years. The divergence of the Indo-European languages to form the immense diversity of Hindi, German, Italian, English, and so on took about 6,000 years (Dixon 1997). How can we imagine what has changed since the emergence of *Homo sapiens* some 200,000 years ago? Or in 5,000,000 years of prior hominid evolution? I claim that the first *Homo sapiens* were language-ready but did not have language in the modern sense. Rather, my hypothesis is that stage S7, the transition from protolanguage to language, is the culmination of manifold discoveries in the history of mankind:

In section 2, I asserted that in much of protolanguage, a complete communicative act involved a unitary utterance, the use of a single symbol formed as a sequence of gestures, whose component gestures – whether manual or vocal – had no independent meaning. Unitary utterances such as "grooflook" or "koomzash" might have encoded quite complex descriptions such as "The alpha male has killed a meat animal and now the tribe has a chance to feast together. Yum, yum!" or commands such as "Take your spear and go around the other side of that animal and we will have a better chance together of being able to kill it." On this view, "protolanguage" grew by adding arbitrary novel unitary utterances to convey *complex but frequently important* situations, and it was a major later discovery en route to language as we now understand it that one could gain expressive power by *fractionating* such utterances into shorter utterances conveying components of the scene or command (cf. Wray 1998; 2000). Put differently, the utterances of prelanguage were more akin to the "calls" of modern primates – such as the "leopard call" of the vervet monkey, which is emitted by a monkey who has seen a leopard and which triggers the appropriate escape behavior in other monkeys – than to sentences as defined in a language like English, but they differed *crucially* from the primate calls in that new utterances could be invented and acquired through learning within a community, rather than emerging only through biological evolution. Thus, the set of such unitary utterances was open, whereas the set of calls was closed.

The following hypothetical but instructive example is similar to examples offered at greater length by Wray (1998; 2000) to suggest how the fractionation of unitary utterances might occur (and see Kirby [2000] for a related computer simulation): Imagine that a tribe has two unitary utterances concerning fire which, by chance, contain similar substrings which become regularized so that for the first time there is a sign for "fire." Now the two original utterances are modified by replacing the similar substrings by the new regularized substring. Eventually, some tribe members regularize the complementary gestures in the first string to get a sign for "burns"; later, others regularize the complementary gestures in the second string to get a sign for "cooks meat." However, because of the arbitrary origin of the sign for "fire," the placement of the gestures that have come to denote "burns" relative to "fire" differs greatly from those for "cooks meat" relative to "fire." It therefore requires a further invention to regularize the placement of the gestures in both utterances – and in the process, words are crystallized at the same time as the protosyntax that combines them. Clearly, such fractionation could apply to protosign as well as to protospeech.

However, fractionation is not the only mechanism that could produce composite structures. For example, a tribe might over the generations develop different signs for "sour apple," "ripe apple," "sour plum," "ripe plum," and so on, but not have signs for "sour" and "ripe" even though the distinction is behaviorally important. Hence, *2n* signs are needed to name *n* kinds of fruit. Occasionally someone will eat a piece of sour fruit by mistake and make a characteristic face and intake of breath when doing so. Eventually, some genius pioneers the innovation of getting a conventionalized variant of this gesture accepted as the sign for "sour" by the community, to be used as a warning before eating the fruit, thus extending the protolanguage.[25] A step towards language is taken when another genius gets people to use the sign for "sour" plus the sign for "ripe X" to replace the sign for "sour X" for each kind X of fruit. This innovation allows new users of the protolanguage to simplify learning fruit names, since now only $n + 1$ names are required for the basic vocabulary, rather than *2n* as before. More to the point, if a new fruit is discovered, only one name need be invented rather than two. I stress that the invention of "sour" is a great discovery in and of itself. It might take hundreds of such discoveries distributed across centuries or more before someone could recognize the commonality across all these constructions and thus invent the precursor of what we would now call adjectives.[26]

The latter example is meant to indicate how a sign for "sour" could be added to the protolanguage vocabulary with no appeal to an underlying "adjective mechanism." Instead, one would posit that the features of language emerged by bricolage (tinkering) which added many features as "patches" to a protolanguage, with general "rules" emerging both consciously and unconsciously only as generalizations could be imposed upon, or discerned in, a population of ad hoc mechanisms. Such generalizations amplified the power of groups of inventions by unifying them to provide expressive tools of greatly extended range. According to this account, there was no sudden transition from unitary utterances to an elaborate language with a rich syntax and compositional semantics; no point at which one could say of a tribe "Until now they used protolanguage but henceforth they use language."

To proceed further, I need to distinguish two "readings" of a case frame like Grasp(Leo, raisin), as an action-object frame and as a verb-argument structure. I chart the transition as follows:

(1) As an *action-object frame*, Grasp(Leo, raisin) represents the perception that Leo is grasping a raisin. Here the action "grasp" involves two "objects," one the "grasper" Leo and the other the "graspee," the "raisin." Clearly the monkey has the perceptual capability to recognize such a situation[27] and enter a brain state that represents it, with that representation distributed across a number of brain regions. Indeed, in introducing principle LR5 (from hierarchical structuring to temporal ordering) I noted that the ability to translate a hierarchical conceptual structure into a temporally ordered structure of actions is apparent whenever an animal takes in the nature of a visual scene and produces appropriate behavior. *But to have such a capability does not entail the ability to communicate in a way that reflects these structures*. It is also crucial to note here the importance of recognition not only of the action (mediated by F5) but also of the object (mediated by IT). Indeed, Figure 2 (the FARS model) showed that the canonical activity of F5 already exhibits a choice between the affordances of an object (mediated by the dorsal stream) that involves the nature of the object (as recognized by IT and elaborated upon in PFC in a process of "action-oriented perception"). In the same way, the activity of mirror neurons does not rest solely upon the parietal recognition (in PF, Fig. 3) of the hand motion and the object's affordances (AIP) but also on the "semantics" of the object as extracted by IT. In the spirit of Figure 2, I suggest that this semantics is relayed via PFC and thence through AIP and PF to F5 to affect there the mirror neurons as well as the canonical neurons.

(2) My suggestion is that at least the immediate hominid precursors of *Homo sapiens* would have been able to perceive a large variety of action-object frames and, for many of these, to form a distinctive gesture or vocalization to appropriately direct the attention of another tribe member, but that the vocalization used would be in general a unitary utterance which need not have involved separate lexical entries for the action or the objects. However, the ability to symbolize more and more situations would have required the creation of a "symbol tool kit" of meaningless elements[28] from which an open-ended class of symbols could be generated.

(3) As a verb-argument structure, Grasp(Leo, raisin) is expressed in English in a sentence such as "Leo grasps the raisin," with "grasps" the verb, and "Leo" and "raisin" the arguments. I hypothesize that stage S7 was grounded in the development of precursors to verb-argument structure using vocalizations that were decomposable into "something like a verb" and two somethings that would be "something like nouns." This is the crucial step in the transition from protolanguage to human language as we know it. Abstract symbols are grounded (but more and more indirectly) in action-oriented perception; members of a community may acquire the use of these new symbols (the crucial distinction here is with the fixed repertoire of primate calls) by imitating their use by others; and, crucially, these symbols can be compounded in novel combinations to communicate about novel situations for which no agreed-upon unitary communicative symbol exists.

Having stressed above that adjectives are not a "natural category," I hasten to add that I do not regard verbs or nouns as natural categories either. What I do assert is that every human language must find a way to express the content of action-object frames. The vast variety of these frames can yield many different forms of expression across human languages. I view linguistic universals as being based on universals of communication that take into account the processing loads of perception and production rather than as universals of autonomous syntax. Hence, in emphasizing verb-argument structures in the form familiar from English, I am opting for economy of exposition rather than further illustration of the diversities of human language. To continue with the bricolage theme, much of "protosyntax" would have developed at first on an ad hoc basis, with variations on a few basic themes, rather than being grounded from the start in broad categories like "noun" or "verb" with general rule-like procedures to combine them in the phonological expression of cognitive form. It might have taken many, many millennia for people to discover syntax and semantics in the sense of gaining immense expressive power by "going recursive" with a relatively limited set of strategies for compounding and marking utterances. As a language emerged, it would come to include mechanisms to express kinship structures and technologies of the tribes, and these cultural products would themselves be expanded by the increased effectiveness of transmission from generation to generation that the growing power of language made possible. Evans (2003) supports this view by surveying a series of linguistic structures in which some syntactic rules must refer to features of the kinship system which are common in Australian aboriginal tribes but are unknown elsewhere. On this basis, we see such linguistic structures as historical products reflecting the impact of various processes of "cultural selection" on emerging structure.

If one starts with unitary utterances, then symbols that correspond to statements like "Take your spear and go around the other side of that animal and we will have a better chance together of being able to kill it" must each be important enough, or occur often enough, for the tribe to agree on a symbol (e.g., arbitrary string of phonemes) and for each one to replace an elaborate pantomime with a conventionalized utterance of protosign or protospeech. Discovering that separate names could be assigned to each actor, object, and action would require many words instead of one to express such an utterance. However, once the number of utterances with overlap reaches a critical level, economies of word learning would accrue from building utterances from "reusable" components (cf. the Wray-Kirby and "sour fruit" scenarios above). Separating verbs from nouns lets one learn m + n + p words (or less if the same noun can fill two roles) to be able to form m*n*p of the most basic utterances. Of course, not all of these combinations will be useful, but the advantage is that new utterances can now be coined "on the fly," rather than each novel event acquiring group mastery of a novel utterance.

Nowak et al. (2000) analyzed conditions under which a population that had two genes – one for unitary utterances and one for fractionated utterances – would converge into a situation in which one gene or the other (and therefore one type of language or the other) would predominate. But I feel that this misses the whole point: (1) It assumes that there is a genetic basis for this alternative, whereas I believe the basis is historical, without requiring genetic change. (2) It postulates that the alternatives already exist. I believe it is necessary to offer a serious analysis of how both unitary and fractionated utterances came to exist, and of the *gradual process* of accumulating changes that led from the predominance of the former to the predominance of the latter. (3) Moreover, it is not a matter of either/or – modern languages have a predominance of fractionated utterances but make wide use of unitary utterances as well.

The spread of these innovations rested on the ability of other humans not only to imitate the new actions and compounds of actions demonstrated by the innovators, but also to do so in a way that related increasingly general classes of symbolic behavior to the classes, events, behaviors, and relationships that they were to represent. Indeed, consideration of the spatial basis for "prepositions" may help show how visuomotor coordination underlies some aspects of language (cf. Talmy 2000), whereas the immense variation in the use of corresponding prepositions even in closely related languages like English and Spanish shows how the basic functionally grounded semantic-syntactic correspondences have been overlaid by a multitude of later innovations and borrowings.

The transition to *Homo sapiens* thus may have involved "language amplification" through increased speech ability coupled with the ability to name certain actions and objects separately, followed by the ability to create a potentially unlimited set of verb-argument structures and the ability to compound those structures in diverse ways. Recognition of hierarchical structure rather than mere sequencing provided the bridge to constituent analysis in language.

8. Towards a neurolinguistics "beyond the mirror"

Most of the stages of our evolutionary story are not to be seen so much as replacing "old" capabilities of the ancestral brain with new ones, but rather, as extending those capabilities by embedding them in an enriched system. I now build on our account of the evolution of the language-ready brain to offer a synchronic account of the "layered capabilities" of the modern adult human brain.

Aboitiz and García (1997) offer a neuroanatomical perspective on the evolutionary origin of the language areas in the human brain by analyzing possible homologies between language areas of the human brain and areas of the monkey

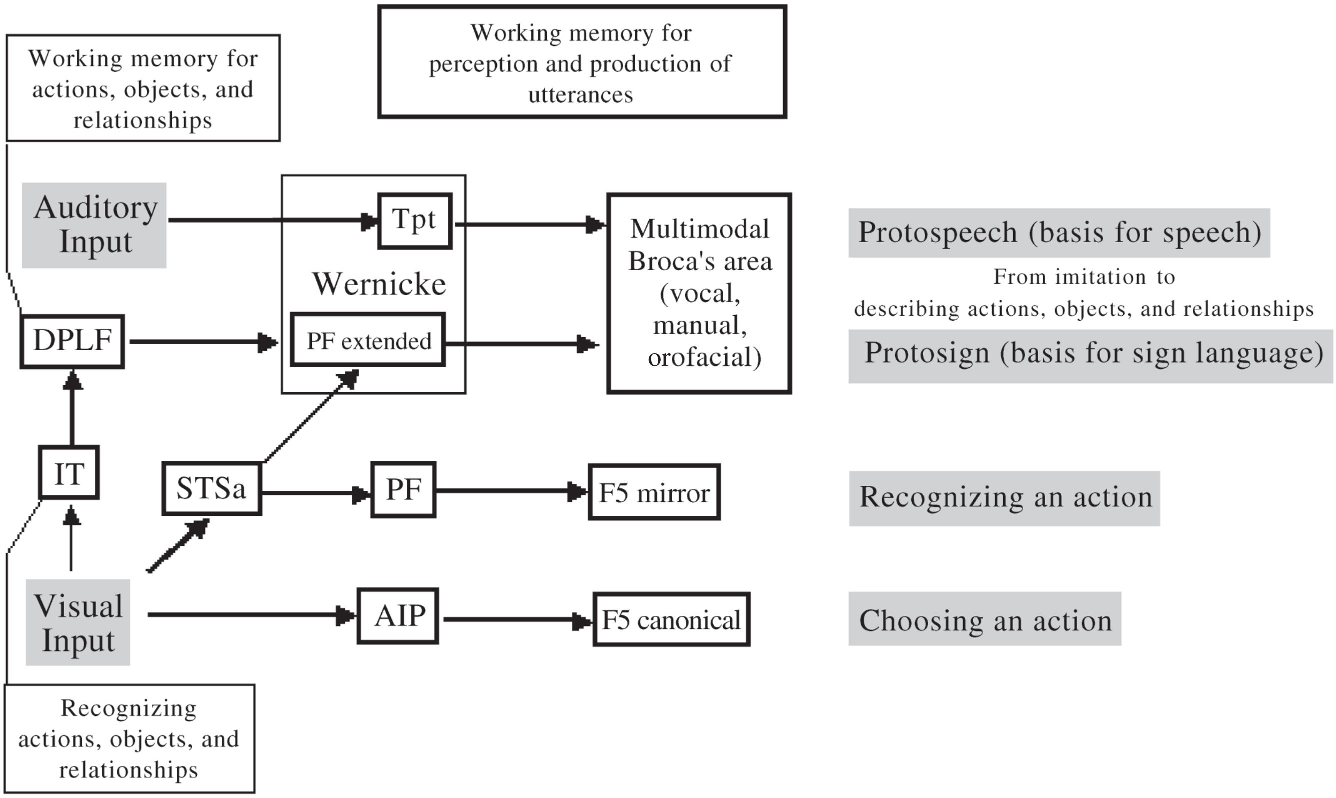


Figure 6. Extending the FARS model to include the mirror system for grasping and the language system evolved "atop" this. Note that this simple figure neither asserts nor denies that the extended mirror system for grasping and the language-supporting system are anatomically separable, nor does it address issues of lateralization. (From Arbib & Bota 2003.)

brain that may offer clues as to the structures of the brains of our ancestors of 20 million years ago. Arbib and Bota (2003) summarize the Aboitiz-García and mirror system hypotheses and summarize other relevant data on homologies between different cortical areas in macaque and human to ground further work on an evolutionary account of the readiness of the human brain for language.

Figure 6 is the diagram Arbib and Bota (2003) used to synthesize lessons about the language mechanisms of the human brain, extending a sketch for a "mirror neurolinguistics" (Arbib 2001b). This figure was designed to *elicit* further modeling; it does not have the status of fully implemented models, such as the FARS and MNS models, whose relation to, and prediction of, empirical results has been probed through computer simulation.

To start our analysis of Figure 6, note that an over-simple analysis of praxis, action understanding, and language production might focus on the following parallel parieto-frontal interactions:

I.	object → AIP → $F5_{canonical}$	praxis
II.	action → PF → $F5_{mirror}$	action understanding
III.	scene → Wernicke's → Broca's	language production

The data on patients A.T. and D.F. reviewed in section 3.1 showed a dissociation between the praxic use of size information (parietal) and the "declaration" of that information either verbally or through pantomime (inferotemporal). D.F. had a lesion allowing signals to flow from V1 towards posterior parietal cortex (PP) but not from V1 to inferotemporal cortex (IT). D.F. could preshape accurately when reaching to grasp an object, even though she was unable to declare, either verbally or in pantomime, the visual parameters that guided the preshape. By contrast, A.T. had a bilateral posterior parietal lesion. A.T. could use her hand to pantomime the size of a cylinder, but could not preshape appropriately when asked to grasp it. This suggests the following scheme:

IV. Parietal "affordances" → preshape
V. IT "perception of object" → pantomime or verbally describe size

That is, one cannot pantomime or verbalize an affordance; but rather one needs a "recognition of the object" (IT) to which attributes can be attributed before one can express them. Recall now the path shown in Figure 2 from IT to AIP, both directly and via PFC. I postulate that similar pathways link IT and PF. I show neither of these pathways in Figure 6, but rather show how this pathway might in the human brain not only take the form needed for praxic actions but also be "reflected" into a pathway that supports the recognition of communicative manual actions. We would then see the "extended PF" of this pathway as functionally integrated with the posterior part of Brodmann's area 22, or area Tpt (temporo-parietal) as defined by Galaburda and Sanides (1980). Indeed, lesion-based views of Wernicke's area may include not only the posterior part of Tpt but also (in whole or in part) areas in the human cortex that correspond to macaque PF (see Arbib & Bota [2003] for further details). In this way, we see Wernicke's area as combining capabilities for recognizing protosign and protospeech to support a language-ready brain that is capable of learning signed languages as readily as spoken languages.

Finally, we note that Arbib and Bota (2003) responded to the analysis of Aboitiz and García (1997) by including a number of working memories crucial to the linkage of visual scene perception, motor planning, and the production and recognition of language. However, they did not provide data on the integration of these diverse working memory systems into their anatomical scheme.

When building upon Figure 6 in future work in neurolinguistics, we need to bear in mind the definition of "complex imitation" as the ability to recognize another's performance as a set of familiar movements and then repeat them, but also to recognize when such a performance combines novel actions that can be approximated by (i.e., more or less crudely be imitated by) variants of actions already in the repertoire. Moreover, in discussing the FARS model in section 3.1, I noted that the interactions shown in Figure 2 are supplemented in the computer implementation of the model by code representing the role of the basal ganglia in administering sequences of actions, and that Bischoff-Grethe et al. (2003) model the possible role of the basal ganglia in interactions with the pre-SMA in sequence learning. Therefore, I agree with Visalberghi and Fragaszy's (2002, p. 495) suggestion that "[mirror] neurons provide a neural substrate for segmenting a stream of action into discrete elements matching those in the observer's repertoire, as Byrne (1999) has suggested in connection with his string-parsing theory of imitation," while adding that the success of complex imitation requires that the appropriate motor system be linked to appropriate working memories (as in Fig. 6) as well as to pre-SMA and basal ganglia (not shown in Fig. 6) to extract and execute the overall structure of the compound action (which may be sequential, or a more general coordinated control program [Arbib 2003]). Lieberman (2002) emphasizes that the roles of Broca's and Wernicke's areas must be seen in relation to larger neocortical and subcortical circuits. He cites data from studies of Broca's aphasia, Parkinson's disease, focal brain damage, and so on, to demonstrate the importance of the basal ganglia in sequencing the elements that constitute a complete motor act, syntactic process, or thought process. Hanakawa et al. (2002) investigated numerical, verbal, and spatial types of nonmotor mental-operation tasks. Parts of the posterior frontal cortex, consistent with the pre-supplementary motor area (pre-SMA) and the rostral part of the dorsolateral premotor cortex (PMdr), were active during all three tasks. They also observed activity in the posterior parietal cortex and cerebellar hemispheres during all three tasks. An fMRI study showed that PMdr activity during the mental-operation tasks was localized in the depths of the superior precentral sulcus, which substantially overlapped the region active during complex finger movements and was located dorsomedial to the presumptive frontal eye fields.

Such papers are part of the rapidly growing literature that relates human brain mechanisms for action recognition, imitation, and language. A full review of such literature is beyond the scope of the target article, but let me first list a number of key articles – Binkofski et al. (1999), Decety et al. (1997), Fadiga et al. (2002), Grezes et al. (1998), Grezes and Decety (2001; 2002), Heiser et al. (2003), Hickok et al. (1998), Iacoboni et al. (1999; 2001), and Floel et al. (2003) – and then briefly describe a few others:

Koski et al. (2002) used fMRI to assess the effect of explicit action goals on neural activity during imitation. Their results support the hypothesis that areas relevant to motor preparation and motor execution are tuned to coding goal-oriented actions and are in keeping with single-cell recordings revealing that neurons in area F5 of the monkey brain represent goal-directed aspects of actions. Grezes et al. (2003) used event-related fMRI to investigate where in the human brain activation can be found that reflects both canonical and mirror neuronal activity. They found activation in the intraparietal and ventral limbs of the precentral sulcus when subjects observed objects and when they executed movements in response to the objects ("canonical neurons"); and activation in the dorsal premotor cortex, the intraparietal cortex, the parietal operculum (SII), and the superior temporal sulcus when subjects observed gestures ("mirror neurons"). Finally, activations in the ventral premotor cortex and inferior frontal gyrus (Brodmann area [BA] 44) were found when subjects imitated gestures and executed movements in response to objects. These results suggest that in the human brain, the ventral limb of the precentral sulcus may form part of the area designated F5 in the macaque monkey. It is possible that area 44 forms an anterior part of F5, though anatomical studies suggest that it may be a transitional area between the premotor and prefrontal cortices.

Manthey et al. (2003) used fMRI to investigate whether paying attention to objects versus movements modulates premotor activation during the observation of actions. Participants were asked to classify presented movies as showing correct actions, erroneous actions, or senseless movements. Erroneous actions were incorrect either with regard to employed objects, or to performed movements. The ventrolateral premotor cortex (vPMC) and the anterior part of the intraparietal sulcus (aIPS) were strongly activated during the observation of actions in humans. Premotor activation was dominantly located within BA 6, and sometimes extended into BA 44. The presentation of object errors and movement errors showed that left premotor areas were more involved in the analysis of objects, whereas right premotor areas were dominant in the analysis of movements. (Since lateralization is not analyzed in this article, such data may be a useful springboard for commentaries.)

To test the hypothesis that action recognition and language production share a common system, Hamzei et al. (2003) combined an action recognition task with a language production task and a grasping movement task. Action recognition-related fMRI activation was observed in the left inferior frontal gyrus and on the border between the inferior frontal gyrus (IFG) and precentral gyrus (PG), the ventral occipito-temporal junction, the superior and inferior parietal cortex, and in the intraparietal sulcus in the left hemisphere. An overlap of activations due to language production, movement execution, and action recognition was found in the parietal cortex, the left inferior frontal gyrus, and the IFG-PG border. The activation peaks of action recognition and verb generation were always different in single subjects, but no consistent spatial relationship was detected, presumably suggesting that action recognition and language production share a common functional architecture, with functional specialization reflecting developmental happenstance.

Several studies provide behavioral evidence supporting the hypothesis that the system involved in observation and preparation of grasp movements partially shares the cortical areas involved in speech production. Gentilucci (2003a) had subjects pronounce either the syllable *ba* or *ga* while observing motor acts of hand grasp directed to objects of two sizes, and found that both lip aperture and voice peak amplitude were greater when the observed hand grasp was directed to the large object. Conversely, Glover and Dixon (2002; see Glover et al. 2004 for related results) presented subjects with objects on which were printed either the word *large* or *small*. An effect of the words on grip aperture was found early in the reach, but this effect declined continuously as the hand approached the target, presumably due to the effect of visual feedback. Gerlach et al. (2002) showed that the left ventral premotor cortex is activated during categorization not only for tools but also for fruits and vegetables and articles of clothing, relative to animals and nonmanipulable man-made objects. Such findings support the notion that certain lexical categories may evolve from action-based knowledge but are difficult to account for should knowledge representations in the brain be truly categorically organized.

Several insights have been gleaned from the study of signed language. Corina et al. (2003) used PET to examine deaf users of ASL as they generated verb signs independently with their right dominant and left nondominant hands (compared to the repetition of noun signs). Nearly identical patterns of left inferior frontal and right cerebellum activity were observed, and these were consistent with patterns that have been reported for spoken languages. Thus, lexical-semantic processing in production relies upon left-hemisphere regions regardless of the modality in which a language is realized, and, in signing, no matter which hand is used. Horwitz et al. (2003) studied the activation of Broca's area during the production of spoken and signed language. They showed that BA45, not BA44, was activated by both speech and signing during the production of language narratives in bilingual subjects (fluent from early childhood in both ASL and English) with the generation of complex movements and sounds as control. Conversely, BA44, not BA45, was activated by the generation of complex articulatory movements of oral-laryngeal or limb musculature. Horwitz et al. therefore conclude that BA45 is the part of Broca's area that is fundamental to the modality-independent aspects of language generation.

Gelfand and Bookheimer (2003), using fMRI, found that the posterior portion of Broca's area responded specifically to sequence manipulation tasks, whereas the left supramarginal gyrus was somewhat more specific to sequencing phoneme segments. These results suggest that the left posterior inferior frontal gyrus responds not to the sound structure of language but rather to sequential operations that may underlie the ability to form words out of dissociable elements.

Much more must be done to take us up the hierarchy from elementary actions to the recognition and generation of novel compounds of such actions. Nonetheless, the above preliminary account strengthens the case that no powerful syntactic mechanisms need have been encoded in the brain of the first *Homo sapiens*. Rather, it was the extension of the imitation-enriched mirror system to support intended communication that enabled human societies, across many millennia of invention and cultural evolution, to achieve human languages in the modern sense.

ACKNOWLEDGMENTS

The early stages of building upon "Language within Our Grasp" (Rizzolatti & Arbib 1998) were conducted during my sabbatical visits in 1999 to the University of Western Australia and the Institute of Human Physiology in Parma, Italy, and my conversations there with Robyn Owens, E. J. Holden, Giacomo Rizzolatti, Morten Christiansen, Giuseppe Cossu, Giuseppe Luppino, Massimo Matelli, Vittorio Gallese, and other colleagues. So many people have offered perceptive comments on various results of that effort (as published in, e.g., Arbib 2001a; 2001b; 2002) that the following list is woefully incomplete – Shannon Casey, Chris Code, Bob Damper, Kerstin Dautenhahn, Barry Gordon, Jim Hurford, Bipin Indurkhya, Chrystopher Nehaniv, and Chris Westbury – but I do hope that all these people (and the *BBS* referees), whether named or not, will realize how much I value their thoughtful comments and that they will see how their suggestions and comments have helped me clarify, correct, and extend my earlier analyses.

Preparation of the present paper was supported in part by a fellowship from the Center for Interdisciplinary Research of the University of Southern California. In particular, this fellowship allowed me to initiate a faculty seminar in September of 2002 at which my ideas have been exposed to intense though friendly scrutiny and placed in the context of the range of fascinating work by the members of the seminar – Amit Almor, Elaine Andersen, Aude Billard, Mihail Bota, Dani Byrd, Vincent Chen, Karen Emmorey, Andrew Gordon, James Gordon, Jack Hawkins, Jerry R. Hobbs, Laurent Itti, Toby Mintz, Stefan Schaal, Craig Stanford, Jean-Roger Vergnaud, Christoph von der Malsburg, Carolee Winstein, Michail Zak, Patricia Zukow-Goldring, and Kie Zuraw.

NOTES

1. Bickerton (1995) views infant language, pidgins, and the "language" taught to apes as *protolanguages* in the sense of a form of communication whose users can only string together a small handful of words at a time with little if any syntax. Bickerton hypothesizes that the protolanguage (in my sense) of *Homo erectus* was a protolanguage in his sense, in which a few words much like those of today's language are uttered a few at a time to convey meaning without the aid of syntax. I do not assume (or agree with) this hypothesis.

2. Today's signed languages are fully expressive human languages with a rich syntax and semantics, and are not to be confused with the posited systems of protosign communication. By the same taken, protospeech is a primitive form of communication based on vocal gestures but without the richness of modern human spoken languages.

3. Since we will be concerned in what follows with sign language as well as spoken language, the "speaker" and "hearer" may be using hand and face gestures rather than vocal gestures for communication.

4. However, I shall offer below the view that early forms of protosign provided a scaffolding for the initial development of protospeech, rather than holding that protosign was "completed" before protospeech was "initiated."

5. I would welcome commentaries on "language-like" aspects of communication in nonprimates, but the present article is purely about changes within the primates that led to the human language-ready brain.

6. It could be objected that monkey calls are not "involuntary communication" because, for example, vervet alarm calls are given usually in the presence of conspecifics who would react to them. However, I would still call this involuntary – this just shows that two conditions, rather than one, are required to trigger the call. This is distinct from the human use of language to conduct a conversation that may have little or no connection to the current situation.

7. When I speak of a "stage" in phylogeny, I do not have in mind an all-or-none switch in the genotype that yields a discontinuous change in the phenotype, but rather the coalescence of a variety of changes that can be characterized as forming a global pattern that may emerge over the course of tens or even hundreds of millennia.

8. Let me stress that complex imitation involves both the recognition of an action as a certain combination of actions and the ability to replicate (something like) that combination. Both skills play a role in the human child's acquisition of language; the latter remains important in the adult's language comprehension.

9. But see note 4 above.

10. The attainment of complex imitation was seen as a crucial stage of the evolution of language readiness in Arbib (2002), but was not listed there as a condition for language readiness. I now see this as a mistake.

11. Unfortunately, space does not permit development of an argument for this controversial claim. Commentaries pro or con the hypothesis will be most welcome.

12. I wonder at times whether properties LR1 through LR7 do indeed support LA1 or whether LA1 should itself be seen as part of the biological equipment of language readiness. I would welcome commentaries in support of either of these alternatives. However, I remain convinced that LR1 through LR7 coupled with LA1 provide all that is needed for a brain to support LA2, LA3, and LA4.

13. The pairs (LR6: Beyond the here-and-now 1; LA3: Beyond the here-and-now 2) and (LR7: Paedomorphy and sociality; LA4: Learnability) do not appear in Table 1 because the rest of the paper will not add to their brief treatment in section 2.2.

14. Figure 2 provides only a partial overview of the model. The full model (see Fagg & Arbib 1998 for more details) includes a number of brain regions, offering schematic models for some and detailed neural-network models for others. The model has been implemented on the computer so that simulations can demonstrate how the activities of different populations vary to explain the linkage between visual affordance and manual grasp.

15. To keep the exposition compact, in what follows I will use without further explanation the abbreviations for the brain regions not yet discussed. The reader wanting to see the abbreviations spelled out, as well as a brief exposition of data related to the hypothesized linkage of schemas to brain structures, is referred to Oztop and Arbib (2002).

16. Estimates for the timetable for hominid evolution (I use here those given by Gamble 1994, see his Fig. 4.2) are 20 million years ago for the divergence of monkeys from the line that led to humans and apes, and 5 million years ago for the divergence of the hominid line from the line that led to modern apes.

17. For more on "chimpanzee culture," see Whiten et al. (2001) and the Chimpanzee Cultures Web site: http://culture.st-and.ac.uk:16080/chimp/, which gives access to an online database that describes the cultural variations in chimpanzee behavior and shows behavior distributions across the sites in Africa where long-term studies of chimpanzees have been conducted in the wild.

18. Recall the observation (Note 8) that both the recognition of an action as a certain combination of actions and the ability to replicate (something like) that combination play a role in the human child's acquisition of language, while the former remains important in the adult's language comprehension. But note, too, that stage S4 only takes us to complex imitation of *praxic* actions; Sections 5 and 6 address the transition to an open system of *communicative* actions.

19. As ahistorical support for this, note that *airplane* is signed in American Sign Language (ASL) with tiny repeated movements of a specific handshape, whereas *fly* is signed by moving the same handshape along an extended trajectory (Supalla & Newport 1978). I say "ahistorical" because such signs are part of a modern human language rather than holdovers from protosign. Nonetheless, they exemplify the mixture of iconicity and convention that, I claim, distinguishes protosign from pantomime.

20. Of course, relatively few Chinese characters are so pictographic in origin. For a fuller account of the integration of semantic and phonetic elements in Chinese characters (and a comparison with Sumerian logograms) see Chapter 3 of Coulmas 2003.

21. Of course, those signs that most clearly resemble pantomimes will be easier for the nonsigner to recognize, just as certain Chinese characters are easier for the novice to recognize. Shannon Casey (personal communication) notes that moving the hands in space to represent actions involving people interacting with people, animals, or other objects is found in signed languages in verbs called "spatial verbs" or "verbs of motion and location." These verbs can be used with handshapes to represent people or objects called "semantic classifiers" and "size and shape specifiers" (Supalla 1986; see p. 196 for a description of these classifiers and p. 211 for figures of them). Hence, to describe giving someone a cup, the ASL signer may either use the standard *give* handshape (palm up with fingertips and thumb-tip touching) or use an open, curved handshape with the fingertips and thumb-tip apart and the palm to the side (as if holding a cup). Similarly, to describe giving someone a thick book, the signer can use a handshape with the palm facing up, fingertips pointing outward and thumb also pointing outward with about an inch of space between the thumb and fingertips (as if holding a book). In her own research Casey (2003) has found that hearing subjects with no knowledge of a signed language do produce gestures resembling classifiers. Stokoe (2001, pp. 188–91) relates the use of shape classifiers in ASL to the use of shape classifiers in spoken Native American languages.

22. Added in proof: Hurford notes that this suggestion was made and discarded prior to publication of Hurford (2004).

23. Such developments and inventions may have occurred very slowly over the course of many (perhaps even thousands) of generations during which expansion of the proto-vocabulary was piecemeal; it may then have been a major turning point in human history when it was realized that symbols could be created *ad libitum* and this realization was passed on to future generations. See also Note 25.

24. Where Corballis focuses on the *FOXP2* gene, Crow (2002a) links lateralization and human speciation to a key mutation which may have speciated on a change in a homologous region of the X and Y chromosomes.

25. I use the word "genius" advisedly. I believe that much work on language evolution has been crippled by the inability to imagine that things we take for granted were in no way a priori obvious, or to see that current generalities were by no means easy to discern in the particularities that they embrace. Consider, for example, that Archimedes (c. 287–212 BCE) had the essential idea of the integral calculus, but it took almost 2,000 years before Newton (1642–1727) and Leibniz (1646–1716) found notations that could express the generality implicit in his specific examples and hence unleash an explosion of mathematical innovation. I contend that language, like mathematics, has evolved culturally by such fits and starts. Note 23.

26. Indeed, adjectives are not the "natural category" they may appear to be. As Dixon (1997, pp. 142 et seq.) observes, there are two kinds of adjective classes across human languages: (1) an open class with hundreds of members (as in English); (2) a small closed class. Languages with small adjective classes are found in every continent except Europe. Igbo, from west Africa, has just eight adjectives: *large* and *small*; *black/dark* and *white/light*; *new* and *old*; and *good* and *bad*. Concepts that refer to physical properties tend to be placed in the verb class (e.g., "the stone heavies") and words referring to human propensities tend to be nouns (e.g., "she has cleverness").

27. Leaving aside the fact that the monkey probably does not know that Leo's name is "Leo."

28. Not all the symbols need be meaningless; some signs of a signed language can be recognized as conventionalized pantomime, and some Chinese characters can be recognized as conventionalized pictures. But we have already noted that relatively few Chinese characters are pictographic in origin. Similarly, many signs have no link to pantomime. As Coulmas (2003) shows us in analyzing writing systems – but the point holds equally well for speech and sign – the mixture of economy of expression and increasing range of expression leads to more and more of a symbol being built up from meaningless components.

Open Peer Commentary

Speech and gesture are mediated by independent systems

Anna M. Barrett,[a] Anne L. Foundas,[b] and
Kenneth M. Heilman[c]

[a]*Department of Neurology and Neural and Behavioral Sciences, Pennsylvania State University College of Medicine, Hershey, PA 17033;* [b]*Department of Psychiatry and Neurology, Tulane University School of Medicine and the Veterans Affairs Medical Center, New Orleans, LA 70115;* [c]*College of Medicine, Gainesville, FL 32610, and North Florida South Georgia Veterans Affairs Medical Center, Gainesville, FL 32610.*
ambarrett@psu.edu foundas@tulane.edu
heilman@neurology.ufl.edu
http://www.som.tulane.edu/neurograd/foundahm.htm#Foundas

Abstract: Arbib suggests that language emerged in direct relation to manual gestural communication, that Broca's area participates in producing and imitating gestures, and that emotional facial expressions contributed to gesture-language coevolution. We discuss functional and structural evidence supporting localization of the neuronal modules controlling limb praxis, speech and language, and emotional communication. Current evidence supports completely independent limb praxis and speech/language systems.

Arbib suggests that language coevolved with a neuroanatomic-behavioral system specialized for manual gesture production and imitation, relying upon a shared anatomic substrate. He suggests that the region adapted for this purpose corresponds to Broca's area in the human brain and area F5 in the monkey. Although this is an interesting hypothesis, there are behavioral and anatomic studies inconsistent with this theory.

First, the target article treats four forelimb gesture classes interchangeably: volitional meaningless movements or movements unrelated to tool use (e.g., reaching or grasping movements), movements standing for objects (emblems, e.g., flapping hands for "bird"), tool use pantomimes (limb praxis), and movements carrying linguistic meaning (signs). A comprehensive discussion of the literature supporting separable systems controlling these movement classes is beyond the scope of this commentary. The term "gesture" here refers to limb praxis unless otherwise noted.

The anatomic evidence is weak regarding Broca's area as a shared neuronal substrate for human gesture and language. Although abnormal skilled learned purposive movements (limb apraxia) may co-occur with speech and language disorders (aphasia), these two conditions are anatomically and functionally dissociable (Kertesz et al. 1984; Liepmann 1920; Papagno et al. 1993; see Heilman & Rothi 2003 for a review). Indeed, in left-handed subjects, brain regions supporting limb praxis may be localized in the right hemisphere, and those supporting speech and language, in the left hemisphere (Heilman et al. 1973; Valenstein & Heilman 1979). Right-handed subjects with limb apraxia and gestural production and imitation deficits are hypothesized to have lost function in brain regions supporting time-space-movement representations. Most have damage to posterior, rather than anterior, cortical areas, usually the inferior parietal lobe (Rothi et al. 1994). A report of an isolated disorder of gesture imitation was associated with posterior, not frontal, cortical injury (Ochipa et al. 1994). Premotor lesions have been associated with limb apraxia, but these lesions were in the supplementary motor area (Watson et al. 1986) or the convexity premotor cortex (superior to Broca's area; Barrett et al. 1998). Lesions in Broca's area can cause oral or buccofacial apraxia and apraxia of speech. However, although limb apraxia is associated with Broca's aphasia (Albert et al. 1981), we could find no reports of patients with limb apraxia for whom there was anatomic evidence of a brain lesion restricted to Broca's area. Patients with nonfluent aphasias (e.g., Broca's aphasia) can acquire simplified signs and gestural emblems (Amerind; not a signed language) as an alternative method of communication (Albert et al. 1981; Skelly 1974) and are preferred candidates over patients with aphasia from posterior brain injury. If coactivity of the speech-language and praxis systems led evolutionarily to refinement of these functional systems, one might expect coactivity to be associated with competent gestural and linguistic performance. However, in normal subjects, spontaneously speaking while gesturing is associated with poorer limb praxis performance (Barrett et al. 2002). Thus, anatomic-behavioral evidence in humans does not support Broca's area as necessary for programming skilled learned forelimb gestures.

Although Arbib posits a relationship between primate orofacial gestures, vocalizations communicating emotional state, and human language, human speech-language functions may be localized quite differently from emotionally communicative functions. In most humans, the left hemisphere may be dominant in the control of vocalization associated with propositional speech, but the right hemisphere often controls vocalization associated with emotional prosody (Heilman et al. 2000; Ross 1981; Tucker et al. 1977), automatic speech (e.g., the Lord's Prayer; Speedie et al. 1993) and singing (Wildgruber et al. 1996). In patients with aphasia after left-hemisphere injury, comprehension and production of affective vocal prosody and emotional facial expressions may also be relatively spared (Barrett et al. 1999; Kanter et al. 1986). This double dissociation argues against left-hemisphere dominance for comprehending, imitating, or producing emotional facial expression or prosody.

Lastly, there is evidence that brain anatomical regions mediating language functions are not truly homologues to the equivalent regions in primates. In humans, the classic Broca's area is not one functional unit but is comprised of two anatomical subregions differing in gross morphology, microscopic cytoarchitectonic features, and functionally (Foundas et al. 1998). The human Broca's area includes the pars triangularis (PTR) and pars opercularis (POP), which may mediate different functions. The PTR is comprised of higher-order heteromodal-association cortex suited to complex cross-modal associations typical of higher-order linguistic functions (e.g., syntax, lexical semantics), whereas the POP is comprised of motor-association cortex suited to articulatory and motor speech functions. In a lesion study, Tonkonogy and Goodglass (1981) reported differences between two patients with delimited Broca subregion lesions. A patient with a lesion restricted to the PTR had linguistic deficits, whereas the patient with a lesion to the POP had disordered articulation and fluency, but preserved higher-order language functions. Using functional MRI to study phonemic versus semantic fluency in the inferior frontal gyrus, Paulesu et al. (1997) also found functional heterogeneity within Broca's area. Whereas phonemic and semantic fluency tasks activated the PTR, only phonemic fluency tasks activated the POP. Hence, the anterior triangular portion of Broca's area (PTR) and the POP may be functionally dissociated, with the PTR functioning more critically in lexical retrieval, and the POP selectively subserving articulatory motor speech functions. Nonhuman primates may not have a homologue to the PTR (Brodmann's area 45). Thus, rather than being a PTR homologue, area F5 in the monkey may represent the human POP (Brodmann's area 44), more directly linked to vocalization.

Arbib discusses the possibility that some forms of limb praxis and speech/language function may draw upon the ability to represent symbolic meaning. Although not all people with limb apraxia or aphasia demonstrate a loss of action semantics or linguistic semantics, and Liepmann (1905) rejected asymbolia as a sole explanation for limb apraxia, some patients with limb apraxia (e.g., conceptual apraxia; Heilman & Rothi 2003; Raymer & Ochipa 1997) demonstrate abnormal action-meaning systems apart from gestural tasks. Whether a separation of action semantics and linguistic semantics in the modern human brain developed over the course of human evolution, however, unfortunately cannot be determined at our present level of knowledge.

Beyond the mirror neuron – the smoke neuron?

Derek Bickerton
Department of Linguistics, University of Hawaii, Honolulu, HI 96822.
derbick@hawaii.rr.com

Abstract: Mirror neurons form a poor basis for Arbib's account of language evolution, failing to explain the creativity that must precede imitation, and requiring capacities (improbable in hominids) for categorizing situations and unambiguously miming them. They also commit Arbib to an implausible holophrastic protolanguage. His model is further vitiated by failure to address the origins of symbolization and the real nature of syntax.

Mirror-neuron theory is the second-latest (*FOXP2* is the latest) in a series of magic-bullet solutions to the problems of language evolution. To his credit, Arbib realizes it could not account for all of language. Unfortunately, his attempts to go beyond it fall far short of adequacy.

Even as a significant component of language, mirror neurons are dubious. There cannot be imitation unless someone has first created something to imitate, and mirror neurons offer no clue as to how totally novel sequences – complex ones, at that – could have been created ab ovo. Moreover, when someone addresses you, you don't just imitate what they said (unless you want to be thought an idiot); you say something equally novel.

Arbib treats as wholly unproblematic both the category "frequently occurring situation" and the capacity of pantomime to represent such situations. Situations, frequent or otherwise, do not come with labels attached; indeed, it is questionable whether any species could isolate "a situation" from the unbroken, ongoing stream of experience unless it already had a language with which to do so. For this task requires abstracting away from a potentially infinite number of irrelevant features – place, weather, time of day, number and identity of participants, and on and on. How, short of mind-reading powers that would leave professional clairvoyants gasping, could our alingual ancestors have known which features seemed relevant to the sender of the message, and which did not?

If Arbib answers "through pantomime," one assumes he has never played charades. Those who have, know that even with the help of a large set of "disambiguating signs" – stereotypic gestures for "film title," "book title," and so on, elaborate routines of finger-counting to provide numbers of words and syllables – participants with all the resources of modern human language and cognition find it often difficult and sometimes impossible to guess what the pantomimer is trying to represent. When what is to be represented is not a monosyllabic word but something as complex as "The alpha male has killed a meat animal and now the tribe has a chance to feast together. Yum, yum!" or "Take your spear and go round the other side of that animal and we will have a better chance of being able to kill it" (Arbib's own examples, sect. 7, para. 2), the likelihood of successful guessing becomes infinitesimally small.

Arbib does see what I pointed out more than a decade ago (Bickerton 1990, pp. 97–98),[1] that any espousal of mirror neurons commits one to a holistic (Wray 1998; 2000) rather than a synthetic protolanguage – one that would have to represent "bird flying" with one symbol, rather than two ("bird" and "flying") as all contemporary languages do (see Bickerton [2003] and especially Tallerman [2004] for discussion). True language is then supposed to develop straightforwardly through the "fractionation" of this protolanguage. Arbib asks us to "imagine that a tribe has two unitary utterances concerning fire which, by chance, contain similar substrings" (sect. 7, para. 3). But won't similar substrings also occur in unitary utterances that have nothing to do with fire? Here he is on the horns of a dilemma. If he thinks they will not, he has smuggled in a ready-made word, and if all "similar substrings" behave similarly, a holistic stage becomes superfluous – all the separate words of a synthetic language are already present, clumsily disguised. If he thinks they will – and given the limited number of possible syllables even in modern languages, they will probably occur more often in sequences that have nothing to do with fire – why should they be taken as meaning "fire" in the rarer cases, and what will similar strings in other contexts be assumed to mean? And even before this dilemma can be addressed, Arbib must specify what would count as "similar enough" and explain why phonetic or gestural similarities would not be eroded by natural change processes long before hominids could correlate them with similarities of meaning. Moreover, to extract a symbol meaning "fire" from a holistic utterance, our ancestors must first have had the semantic concept of fire, and it becomes wholly unclear why, instead of going the roundabout holistic route, they could not immediately have given that concept a (signed or spoken) label. Real-world objects can be ostensively defined; events and situations cannot.

Two substantive issues lie at the heart of language evolution: how symbolism emerged, and how syntax emerged. No treatment that fails to deal with both can be taken seriously. Indeed, symbolism (as distinct from iconic or indexical reference, distinctions that Arbib nowhere makes) has seemed to some (e.g., Deacon 1997) to be the Rubicon between our species and others. Arbib mentions it several times, hypothesizing it as a "support" for protolanguage and noting the necessity for its "increasing sophistication" as true language emerges. But at no point does he even acknowledge the problem of how such an evolutionary novelty could have developed.

Syntax makes an even better candidate for a human apomorphy, since even with explicit instruction our nearest relatives fail to acquire the rudiments of it (Givon 2004). Arbib's dismissal of syntax as a "historical phenomenon" makes such uniqueness hard to explain. According to him, "Words as we know them then coevolved culturally with syntax through fractionation" (sect. 2, para. 2). Even if syntax meant only the most frequent word-order in simple affirmative sentences, this claim might be tricky to defend. In fact, syntax depends on a wide variety of relationships within complex hierarchical structures. Where do these structures and relationships come from? Arbib, ignoring the half-century of linguistic research that has revealed (if not explained) them, remains silent on this.

Arbib's treatment claims to go "beyond the mirror." However, what he offers is only a smoke-and-mirrors version of language evolution, one in which all the real issues are obscured. His flowcharts and neurological references may look impressive, but they tell us nothing about the central problems of the field.

NOTE

[1] It is surely worth reminding readers that all the features of mirror neurons (except for their catchy title) were described by David Perrett and his associates (Perrett et al. 1982; 1985) more than two decades ago – a fact seldom acknowledged in contemporary accounts, including Arbib's.

The evolutionary link between mirror neurons and imitation: An evolutionary adaptive agents model

Elhanan Borenstein[a] and Eytan Ruppin[a,b]

[a]*School of Computer Science, Tel Aviv University, Tel-Aviv 69978, Israel;*
[b]*School of Medicine, Tel Aviv University, Tel-Aviv 69978, Israel.*
borens@post.tau.ac.il http://www.cs.tau.ac.il/~borens
ruppin@post.tau.ac.il http://www.cs.tau.ac.il/~ruppin

Abstract: This commentary validates the fundamental evolutionary interconnection between the emergence of imitation and the mirror system. We present a novel computational framework for studying the evolutionary origins of imitative behavior and examining the emerging underlying mechanisms. Evolutionary adaptive agents that evolved in this framework demonstrate the emergence of neural "mirror" mechanisms analogous to those found in biological systems.

Uncovering the evolutionary origins of neural mechanisms is bound to be a difficult task; fossil records or even genomic data can provide very little help. Hence, the author of the target article should be commended for laying out a comprehensive and thorough theory for the evolution of imitation and language. In particular, in considering the first stages in the evolution of language, Arbib argues that the mirror system initially evolved to provide a visual feedback on one's own action, bestowing also the ability to understand the actions of others (stage S2), and that further evolution was required for this system to support the copying of actions and eventually imitation (stages S3 and S4). However, the functional link between the mirror system and the capacity to imitate, although compelling, has not yet been demonstrated clearly. We wish to demonstrate that the existence of a mirror system, capable of matching the actions of self to observed actions of others, is fundamentally linked to imitative behavior and that, in fact, the evolution of imitation promotes the emergence of neural mirroring.

Neurally driven evolutionary adaptive agents (Ruppin 2002) form an appealing and intuitive approach for studying and obtaining insights into the evolutionary origins of the mirror system. These agents, controlled by an artificial neural-network "brain," inhabit an artificial environment and are evaluated according to their success in performing a certain task. The agents' neurocontrollers evolve via genetic algorithms that encapsulate some of the essential characteristics of natural evolution (e.g., inheritance, variation, and selection).

We have recently presented such a novel computational model for studying the emergence of imitative behavior and the mirror system (Borenstein & Ruppin 2004; 2005). In contradistinction to previous engineering-based approaches that explicitly incorporate biologically inspired models of imitation (Billard 2000; Demiris & Hayes 2002; Demiris & Johnson 2003; Marom et al. 2002; Oztop & Arbib 2002), we employ an evolutionary framework and examine the mechanism that evolved to support imitation. Because it is an emerging mechanism (rather than an engineered one), we believe it is likely to share the same fundamental principles driving natural systems.

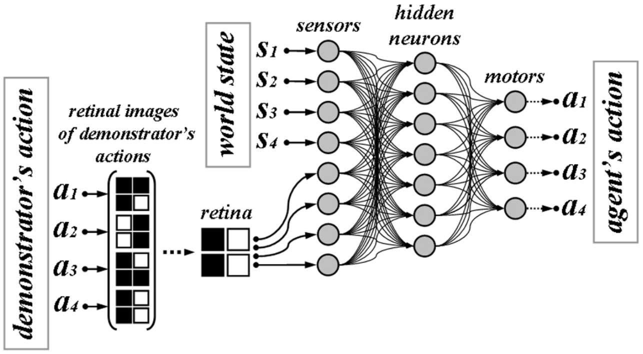


Figure 1 (Borenstein & Ruppin). The agent's sensorimotor system and neurocontroller. The sensory input is binary and includes the current world state and a retinal "image" of the demonstrator's action (when visible). The retinal image for each possible demonstrator's action and a retinal input example are illustrated. The motor output determines which actions are executed by the agent. The network synapses are adaptive, and their connection strength may change during life according to the specified learning rules.

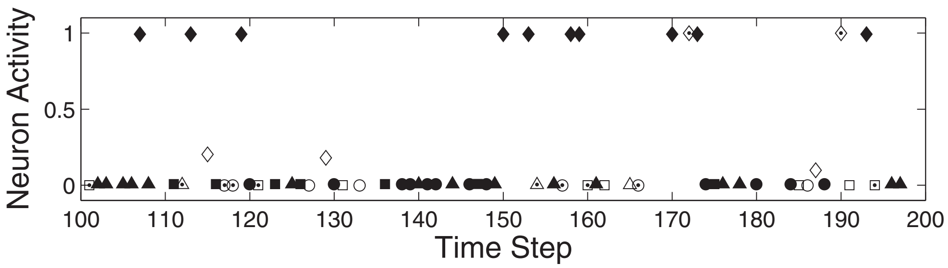
In our model, a population of agents evolves to perform specific actions successfully according to certain environmental cues. Each agent's controller is an adaptive neural network, wherein synaptic weights can vary over time according to various Hebbian learning rules. The genome of these agents thus encodes not only the initial synaptic weights but also the specific learning rule and learning rate that govern the dynamics of each synapse (Floreano & Urzelai 2000). Agents are placed in a changing environment that can take one of several "world states" and should learn to perform the appropriate action in each world state. However, the mapping between the possible world states and appropriate actions is randomly selected anew in the beginning of the agent's life, preventing a successful behavior from becoming genetically determined. Agents can infer the appropriate state-action mapping only from an occasional retinal-sensory input of a demonstrator, successfully performing the appropriate action in each world state (Fig. 1). These settings promote the emergence of an imitation-based learning strategy, although no such strategy is explicitly introduced into the model.

Applying this model, we successfully developed evolutionary adaptive agents capable of learning by imitation. After only a few demonstrations, agents successfully master the behavioral task, regularly executing the proper action in each world state. Moreover, examining the dynamics of the neural mechanisms that have emerged, we found that many of these agents embody a neural mirroring device analogous to that found in biological systems. That is, certain neurons in the network's hidden layer are each associated with a certain action and discharge only when this action is either executed by the agent or observed (Fig. 2). Further analysis of these networks reveals complex dynamics, incorporating both pre-wired perceptual-motor coupling and learned state-action associations, to accomplish the required task.

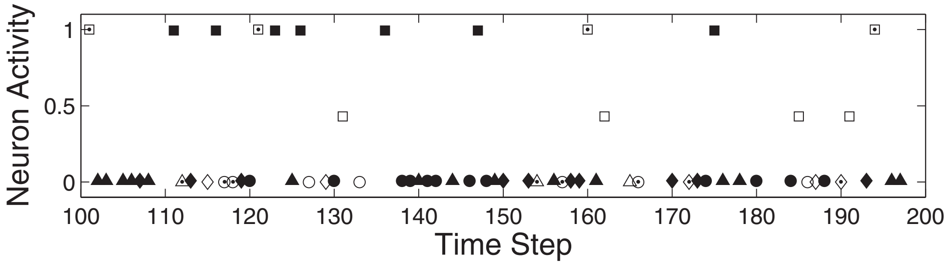
This framework provides a fully accessible, yet biologically plausible, distilled model for imitation and can serve as a vehicle to study the mechanisms that underlie imitation in biological systems. In particular, this simple model demonstrates the crucial role of the mirror system in imitative behavior; in our model, mirror neurons' emergence is derived solely from the necessity to imitate observed actions. These findings validate the strong link between the capacity to imitate and the ability to match observed and executed actions and thus support Arbib's claim for the functional links between the mirror system and imitation. However, whereas Arbib hypothesizes that the evolution of the mirror system preceded the evolution of imitation, this model additionally suggests an alternative possible evolutionary route, grounding the emergence of mirror neurons in the evolution of imitative behavior. Evidently, at least in this simple evolutionary framework, neural mirroring can coevolve in parallel with the evolution of imitation. We believe that evolutionary adaptive agents models, such as the one described above, form a promising test bed for studying the evolution of various neural mechanisms that underlie complex cognitive behaviors. Further research of artificially evolving systems can shed new light on some of the key issues concerning the evolution of perception, imitation, and language.

ACKNOWLEDGMENT
Elhanan Borenstein is supported by the Yeshaya Horowitz Association through the Center for Complexity Science.

(a) Neuron h_4



(b) Neuron h_5



(c) Neuron h_6

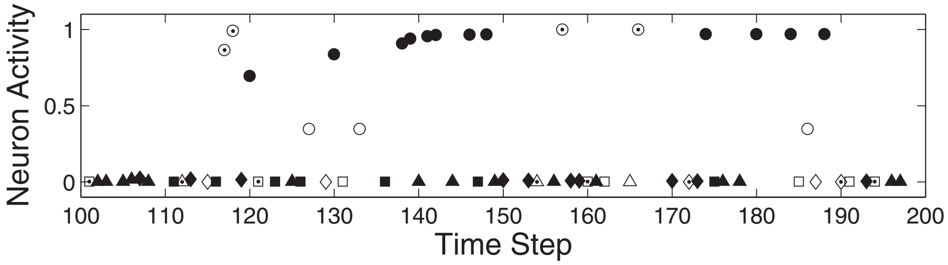


Figure 2 (Borenstein & Ruppin). The activation level of three hidden neurons in a specific successful agent during time steps 100 to 200. Circles, squares, diamonds, and triangles represent the four possible actions in the repertoire. An empty shape indicates that the action was only observed but not executed, a filled shape indicates that the action was executed by the agent (stimulated by a visible world state) but not observed, and a dotted shape indicates time steps in which the action was both observed and executed. Evidently, each of these neurons is associated with one specific action and discharges whenever this action is observed or executed.

Sharpening Occam's razor: Is there need for a hand-signing stage prior to vocal communication?

Conrado Bosman, Vladimir López, and Francisco Aboitiz

Departimento Psiquiatría, Facultad de Medicina, Pontificia Universidad Católica de Chile, Casilla 114-D Santiago 1, Chile. **cbosman@med.puc.cl vlopezh@puc.cl faboitiz@puc.cl http://www.neuro.cl**

Abstract: We commend Arbib for his original proposal that a mirror neuron system may have participated in language origins. However, in our view he proposes a complex evolutionary scenario that could be more parsimonious. We see no necessity to propose a hand-based signing stage as ancestral to vocal communication. The prefrontal system involved in human speech may have its precursors in the monkey's inferior frontal cortical domain, which is responsive to vocalizations and is related to laryngeal control.

In the target article, Arbib extends his earlier hypothesis about the role of mirror neurons for grasping in the motor control of language (Rizzolatti & Arbib 1998), to a more detailed and fine-grained scenario for language evolution. We agree with and celebrate the main proposals that a mirror neuron system has had a fundamental role in the evolution of human communication and that imitation was important in prelinguistic evolution. We also agree that there has probably been an important vocal-gestural interaction in the evolution of communication. In these and other aspects, our viewpoints complement each other (Aboitiz & García 1997). We proposed that language networks originated as a specialization from ancestral working memory networks involved in vocal communication, and Figure 6 of the target article is a good attempt to synthesize both hypotheses. However, we are not so sure yet about the claim that gestural language was a precursor for vocal communication, for several reasons:

First, phylogenetic evidence indicates that in nonhuman primates, vocal communication transmits external meaning (i.e., about events in the world) and is more diverse than gestural communication (Acardi 2003; Leavens 2003; Seyfarth & Cheney 2003a). Second, there is evidence suggesting that the control of vocalizations in the monkey could be partly carried out by cortical areas close to F5 and does not depend exclusively on the anterior cingulate cortex. If this is so, the neural precursor for language would not need to be sought in a hand-based coordination system. For example, in the monkey there is an important overlap between area F5 and the cortical larynx representation (Jürgens 2003). Electrical stimulation of this area can elicit vocal fold movements (Hast et al. 1974), and cortical lesions in the supplementary motor area can significantly reduce the total number of vocalizations emitted by monkeys (Gemba et al. 1997; Kirzinger & Jürgens 1982). Furthermore, Romanski and Goldman-Rakic (2002) recently described, in Brodmann areas 12 and 45 of the monkey, neurons that respond strongly to vocalizations.

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 develop 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 protospeech, it is not clear why protosign simply did not evolve into sign language. The evolutionary advantage of language would already be within the grasp of early humans.

Biological evolution of cognition and culture: Off Arbib's mirror-neuron system stage?

Horacio Fabrega, Jr.
Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA 15213.
hfabregajr@adelphia.net

Abstract: Arbib offers a comprehensive, elegant formulation of brain/language evolution; with significant implications for social as well as biological sciences. Important psychological antecedents and later correlates are presupposed; their conceptual enrichment through protosign and protospeech is abbreviated in favor of practical communication. What culture "is" and whether protosign and protospeech involve a protoculture are not considered. Arbib also avoids dealing with the question of evolution of mind, consciousness, and self.

Is the mirror-neuron system (MNS) purely for grasping a basis for or a consequence of social communication and organization? Arbib suggests that even monkey MNS (involving praxis and vocalization) contains the seeds of or serves "real" communication functions, as does simple imitation (but how?), with respect to social cooperation and ecological problem-solving. Such functions are easier to visualize for emotional and facial gestures than for grasping per se (on which he places most emphasis).

Arbib's formulation of what a pantomime sequence might communicate presupposes enormous cognitive capacities. Much of social cognition, conscious awareness of self and situation, and goal-setting appear already resonant in the brain before pantomime. Some have attributed self-consciousness and the "aboutness relationship" to language (Macphail 2000), but Arbib posits that the reverse occurs.

In Arbib's Table 1, cognitive functions (LR5) are said to precede all of language readiness: This involves a primate being able to take in, decompose, and order a complex perceptual scene as per an action. Yet how this capacity blends into LR1–LR4 is covered mainly in brain terms, with natural selection (behavioral) factors minimized. It is unclear to what extent the idea that much of cognition *precedes and gets recruited* into language readiness differs from formulations of others who cover related topics and whose work is not discussed in detail, such as Deacon (1997), Greenfield (1991), Jackendoff's (1983) and Wilkins and Wakefield's (1995) conceptual structure, the latter's POT (parieto-occipito-temporo junction), McNeilage's (1998) syllabification, and metacommunication and autoneoesis (Suddendorf 1999; Suddendorf & Corballis 1997; Wheeler et al. 1997).

The biological line between LR and L (language) is left open: How much of the protosign/protospeech spiral is enough? Arbib promotes a slow, gradual evolution of LR and L as per communication but handles these as purely in analytical terms, as arbitrarily discontinuous. Despite much work on human speciation events (Crow 2002b), Arbib seems against it. He is vague on "what of" and "how much of" spiraling establishes speciation, the identity of *Homo sapiens*. Is the latter "merely" a cultural event? Arbib suggests that a member of Homo erectus has the capacity to mentally use and associate symbols that are arbitrary and abstract (i.e., showing considerable culture and cognition) yet is able to produce only simple, unitary utterances (showing comparatively little language). This renders ambiguous exactly what marks speciation: Does it come "only" when full language is invented or does is require more cognition and culture (and how much more?) made possible by invention of language as we know it? Behavioral implications of the cognitive/brain jump between simple imitation and complex imitation are also not clearly spelled out. Expressing relationships (compositionally) is said to come later, yet is analogous capacity not perceptually, cognitively inherent even before complex imitation? Exactly how LR differs from simple imitation behaviorally and in terms of the brain is not clear.

Arbib relies on Tomasello's (1999a) idea about the biological capacity for intentional communication and social learning/culture, all inherent in *Homo sapiens* (i.e., with language), yet includes intended communication as part of protolanguage and hence "prehuman." Shared attributes of awareness of self as agent/sender and conspecific as receiver, adding parity and symbolization to this amalgam, are implied. Does this mean that *Homo sapiens*' language-ready brain already enabled self-awareness, self/other differentiation, and social cognition well before its members could actually "do" language? Arbib suggests that culture involves late happenings (Pfeiffer 1982). Does not something like protoculture (Hallowell 1960) accompany LR as the "spiral" begins and gets under way? Arbib also suggests that much of the protosign/protospeech had a learned (cultural?) basis. Behavioral and cognitive implications of Arbib's Language Readiness construct (LR) are abstract and unclear. It (LR) appears to incorporate ordinary ecologic, executive cognition as well as social cognition, and it is unclear how language fits in these. Much of social and executive cognition is collapsed into, seems entailed by, LR schema. What exactly language adds to aspects of self-awareness/consciousness and social cognition is not clear. Fundamental questions of the relation between language and thought are simply bypassed (Carruthers & Boucher 1998).

It is a challenging and very controversial idea that language is a purely cultural achievement – that it was invented and then perfected and remained as a cultural innovation ready for an infant born into a language/culture community to just learn naturally. It is difficult to understand how the articulatory, phonological equipment for language evolved entirely during the pre–*Homo sapiens* LR phase; complexities of speech production seem in excess of what the protosign/protospeech spiraling entails, unless one includes more of language within LR.

Arbib's discussion of LA3 in section 2.2 is stunning: if one removed syntax (how much of it? Arbib mentions only time ordering and numbering system) one would still have language rather than a protolanguage (Bickerton 1995). Arbib does elaborate on this as per time travel but relates it to a whole array of brain/cognition features that support LR6. Is time travel inherent in protolanguage but only used through language? He also suggests that language involves the capacity to exploit these cognitive structures for communication purposes, suggesting that emergence/design of cognitive structures did not have a communicative basis. Did protosign/protospeech spiraling merely have communicational basic functions? This relates to complex language/thought questions which Arbib bypasses.

Why MNS may not have involved vocality along with praxic/gestural features from the start without necessitating a detour of gesture alone is not clear. What brain/genetic conditions were "not in place" that precluded the use of vocality along with manual gesture and that only later made it possible? Arbib's two answers to this conundrum are not entirely persuasive.

The conventionalized gestures used to disambiguate pantomime constitute a major transition into protosign, involving a dissociation between the mirror production system and the recognition system, but this is dealt with by (merely) introducing the hypothesis of intended communication, bypassing problems discussed earlier.

Epilogue: The mirror system as a framework for the evolution of culture. Intellectual quandaries hover over the evolutionary, brain, and social sciences: the nature of consciousness, self-consciousness, psychological experience, cultural knowledge, and selfhood. To understand all of these in terms of brain function, and to bring into this intellectual theater their human evolutionary basis, makes for a very beclouded stage. Many researchers have glided over such questions (Damasio 1987; D'Andrade 1999; Ingold 1996; Wierzbicka 1992; 1993). Some have addressed them in piecemeal fashion (Barkow 1987; Bickerton 1995; Geertz 1973;

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 areas 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.

Although the view of Falk puts language emergence in a continuum that is closer to primate vocal communication than the framework of Michael Arbib, both models involve a progression atop the systems already preexisting in nonhuman primates. Arbib's work gives the first detailed account of putative evolutionary stages in the emergence of human language from a cognitive viewpoint. It therefore could be used as a framework to test specific links between cognitive human language and communicative human language emergence hypotheses, such as the one recently proposed by Falk.

Auditory object processing and primate biological evolution

Barry Horwitz,[a] Fatima T. Husain,[a] and Frank H. Guenther[b]

[a]*Brain Imaging and Modeling Section, National Institute on Deafness and Other Communications Disorders, National Institutes of Health, Bethesda, MD 20892;* [b]*Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215.* **horwitz@helix.nih.gov husainf@nidcd.nih.gov http://www.nidcd.nih.gov/research/scientists/horwitzb.asp guenther@cns.bu.edu http://www.cns.bu.edu/~guenther/**

Abstract: This commentary focuses on the importance of auditory object processing for producing and comprehending human language, the relative lack of development of this capability in nonhuman primates, and the consequent need for hominid neurobiological evolution to enhance this capability in making the transition from protosign to protospeech to language.

The target article by Arbib provides a cogent but highly speculative proposal concerning the crucial steps in recent primate evolution that led to the development of human language. Generally, much of what Arbib proposes concerning the transition from the mirror neuron system to protosign seems plausible, and he makes numerous points that are important when thinking about language evolution. We especially applaud his use of neural modeling to implement specific hypotheses about the neural mechanisms mediating the mirror neuron system. We also think his discussion in section 6 of the necessity to use protosign as scaffolding upon which to ground symbolic auditory gestures in protospeech is a significant insight. However, the relatively brief attention Arbib devotes to the perception side of language, and specifically to the auditory aspects of this perception, seems to us to be a critical oversight. The explicit assumption that protosign developed before protospeech, reinforced by the existence of sign language as a fully developed language, allows Arbib (and others) to ignore some of the crucial features that both the productive and receptive aspects of speech require in terms of a newly evolved neurobiological architecture.

One aspect of auditory processing that merits attention, but is not examined by Arbib, has to do with auditory object processing. By auditory object, we mean a delimited acoustic pattern that is subject to figure-ground separation (Kubovy & Van Valkenburg 2001). Humans are interested in a huge number of such objects (in the form of words, melodic fragments, important environmental sounds), perhaps numbering on the order of 10^5 in an individual. However, it is difficult to train monkeys on auditory object tasks, and the number of auditory objects that interest them, compared to visual objects, seems small, numbering perhaps in the hundreds (e.g., some species-specific calls, some important environmental sounds). For example, Mishkin and collaborators (Fritz et al. 1999; Saunders et al. 1998) have showed that monkeys with lesions in the medial temporal lobe (i.e., entorhinal and perirhinal cortex) are impaired relative to unlesioned monkeys in their ability to perform correctly a visual delayed match-to-sample task when the delay period is long, whereas both lesioned and unlesioned monkeys are equally unable to perform such a task using auditory stimuli.

These results implicate differences in monkeys between vision and audition in the use of long-term memory for objects. Our view is that a significant change occurred in biological evolution allowing hominids to develop the ability to discriminate auditory objects, to categorize them, to retain them in long-term memory, to manipulate them in working memory, and to relate them to articulatory gestures. It is only the last of these features that Arbib discusses. In our view, the neural basis of auditory object processing will prove to be central to understanding human language evolution. We have begun a systematic approach combining neural modeling with neurophysiological and functional brain imaging data to explore the neural substrates for this type of processing (Husain et al. 2004).

Concerning language production, Arbib's model of the mirror-neuron system (MNS) may require considerable modification, especially when the focus shifts to the auditory modality. For instance, there is no treatment of babbling, which occurs in the development of both spoken and sign languages (Petitto & Marentette 1991). Underscoring the importance of auditory processing in human evolution, hearing-impaired infants exhibit vocal babbling that declines with time (Stoel-Gammon & Otomo 1986).

However, there has been work in developing biologically plausible models of speech acquisition and production. In one such model (Guenther 1995), a role for the MNS in learning motor commands for producing speech sounds has been posited. Prior to developing the ability to generate speech sounds, an infant must learn what sounds to produce by processing sound examples from the native language. That is, he or she must learn an auditory target for each native language sound. This occurs in the model via a MNS involving speech sound-map cells hypothesized to correspond to mirror neurons (Guenther & Ghosh 2003). Only after learning this auditory target can the model learn the appropriate motor commands for producing the sound via a combination of feedback and feed-forward control subsystems. After the commands are learned, the same speech sound-map cell can be activated to read out the motor commands for producing the sound. In this way, mirror neurons in the model play an important role in both the acquisition of speaking skills and in subsequent speech production in the tuned system. This role of mirror neurons in development of new motor skills differs from Arbib's MNS model, which "makes the crucial assumption that the grasps that the mirror system comes to recognize are already in the (monkey or human) infant's repertoire" (sect. 3.2, para. 7).

Our efforts to comprehend the biological basis of language evolution will, by necessity, depend on understanding the neural substrates for human language processing, which in turn will rely heavily on comparative analyses with nonhuman primate neurobiology. All these points are found in Arbib's target article. A crucial aspect, which Arbib invokes, is the necessary reliance on neurobiologically realistic neural modeling to generate actual implementations of neurally based hypotheses that can be tested by comparing simulated data to human and nonhuman primate experimental data (Horwitz 2005). It seems to us that the fact that humans use audition as the primary medium for language expression means that auditory neurobiology is a crucial component that must be incorporated into hypotheses about how we must go beyond the mirror-neuron system.

On the neural grounding for metaphor and projection

Bipin Indurkhya

International Institute of Information Technology, Hyderabad 500 019, India. **bipin@iiit.net**

Abstract: Focusing on the mirror system and imitation, I examine the role of metaphor and projection in evolutionary neurolinguistics. I suggest that the key to language evolution in hominid might be an ability to project one's thoughts and feelings onto another agent or object, to see and feel things from another perspective, and to be able to empathize with another agent.

With regard to the evolutionary framework for neurolinguistics spelled out in Arbib's article, I would like to focus on the role of metaphor and projection therein. In particular, I am interested in the implications of Arbib's framework for the thesis "all knowledge (or language) is metaphorical." It should be clarified from the outset that this thesis is sometimes misconstrued to suggest that the literal or conventional does not exist – a suggestion that is trivially refuted. However, the sense in which I take it here is based on a well-known phenomenon that a novel metaphor sometimes becomes conventional through repeated use, and may even turn into polysemy; and the claim is that all that is conventional and literal now must have been metaphorical once (Indurkhya 1994). Furthermore, I take the viewpoint that the key mechanism underlying metaphor, especially creative metaphor, is that of projection, which carves out a new ontology for the target of the metaphor (Indurkhya 1992; 1998). This mechanism can be best explained as projecting a structure onto a stimulus, as in gestalt interaction, and is to be contrasted with the mapping-based approaches to metaphor, which require a pre-existing ontology for mapping. For example, in the context of Arbib's article, it is the projection mechanism that determines what constitutes objects and actions when a monkey watches a raisin being grasped by another monkey or by a pair of pliers.

There are two particular places in the evolutionary account articulated by Arbib where a projection step is implicit, and I shall zoom in on them in turn to raise some open issues. The first of these concerns the mirror neurons (sect. 3.2). Now, certain mirror neurons are known to fire when a monkey observes another monkey performing a particular grasping action but not when the grasp is being performed with a tool. This suggests a predisposition towards the ontology of a biological effector. The interesting question here is: How much variation can be introduced in the effector so that it is still acceptable to the mirror neuron. Does a robot arm trigger the mirror neuron? What about a hairy robot arm?

Similar remarks can be made with respect to the learning effect in mirror neurons. When a monkey first sees a raisin being grasped with a pair of pliers, then his mirror neurons do not fire. However, after many such experiences, the monkey's mirror neurons encoding precision grip start firing when he sees a raisin being grasped with pliers. This shows a predisposition towards the ontology of the object raisin and the effect of grip on it, as it is not the physical appearance of the effector but its effect on the object that matters. Again we may ask how much variation is possible in the object and the kind of grip before the mirror system fails to learn. For example, after the mirror neurons learn to fire on seeing a raisin being grasped with pliers, do they also fire when tweezers are used? Or, does the tweezers grasp have to be learned all over again?

These issues become more prominent when we consider imitation (sect. 4). In the literature, a wide range of animal behaviors are classified as imitation (Caldwell & Whiten 2002; Zentall & Akins 2001), and true imitation is distinguished from imprinting, stimulus enhancement, emulation learning, and so on. However, even in imitating a single action, one has to decide what aspect of the situation to imitate, as any situation has many possible aspects; and how to imitate, as the imitating agent has to interpret the situation from its point of view – it may not have the same effectors, access to the same objects, and so on – and project the observed action into its own action repertoire (Alissandrakis et al. 2002; Hofstadter 1995). In this respect, studies on the behavior of animals that imitate a non-conspecific model, such as bottlenose dolphins or parrots imitating a human model (or a bottlenose dolphin imitating a parrot?) are most illuminating. (See, e.g., Bauer & Johnson 1994; Kuczaj et al. 1998; Moore 1992.) In Arbib's framework, a distinction is made between simple and complex imitation to explain where humans diverge from monkeys, and a projection-like mechanism is posited for complex imitation (sect. 2.1: LR1; also sect. 5). But I would like to suggest that even simple imitation could invoke projection, and the crux of the distinction between humans and other animals might lie in the ability to interpret a wider variety of actions and situations, and to project oneself into those situations to imitate them in a number of ways.

Empathy – being able to put oneself into another's shoes and to project one's thoughts and feelings into another person, animal, or object – is often considered a hallmark of being human. Indeed, one of the ideals of robotics research is to emulate this essentially human trait in robots. (See, e.g., Breazeal et al. 2005; Kozima et al. 2003. This is also the theme of the classic Philip K. Dick story "Do Androids Dream of Electric Sheep?" upon which the popular film *Blade Runner* was based.) A glimpse of the key role played by empathy in human cognition is provided by a study by Holstein (1970), in which children were given projection tasks such as being asked to imagine being a doorknob or a rock, and to describe one's thoughts and feelings in order to stimulate their creativity. In a very recent study, it was found that when participants hid one of their hands and a rubber hand was placed in front of them to make it look like their own hand, it took them only 11 seconds to project their feelings onto the rubber hand as if it were their own, down to the neural level: when the rubber hand was stroked by a brush, the somatosensory area in the participants' brain corresponding to their hand was stimulated (Ehrsson et al. 2004). One wonders if monkeys and other animals are capable of projecting their selves into other animals or other objects to this degree, and if the divergent point of hominid evolution might not be found therein.

Listen to my actions!

Jonas T. Kaplan and Marco Iacoboni

UCLA Brain Mapping Center, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095. **jonask@ucla.edu**
iacoboni@loni.ucla.edu **http://www.jonaskaplan.com**

Abstract: We believe that an account of the role of mirror neurons in language evolution should involve a greater emphasis on the auditory properties of these neurons. Mirror neurons in premotor cortex which respond to the visual and auditory consequences of actions allow for a modality-independent and agent-independent coding of actions, which may have been important for the emergence of language.

We agree with Arbib that the mirror property of some motor neurons most probably played an important role in the evolution of language. These neurons allow us to bridge the gap between two minds, between perception and action. As strong evidence for the role of mirror-like mechanisms in language, we have recently demonstrated with functional magnetic resonance imaging (fMRI) that a human cortical area encompassing primary motor and premotor cortex involved in the production of phonemes is also active during the perception of those same phonemes (Wilson et al. 2004). This suggests that motor areas are recruited in speech perception in a process of auditory-to-articulatory transformation that accesses a phonetic code with motor properties (Liberman et al. 1967).

However, we direct our commentary mostly at what Arbib calls the transition from protosign to protospeech. In Arbib's account, a system of iconic manual gestures evolved from a mirror system of action recognition, and then somehow transitioned to a vocal-based language. Mention is made of the so-called audiovisual mirror neurons, which respond to the sound of an action as well as during the production of that action (Kohler et al. 2002). The role of these neurons in the evolution of language deserves more attention.

Arbib argues that arbitrary signs first evolved in gesture, which was more amenable to iconic representation, and that this protosign provided the "scaffolding" for vocal-based abstractions. We suggest that rather than being added on later, the auditory responsiveness of premotor neurons may have played a more central role in the development of abstract representations. The audiovisual 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

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).

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 terms of the motor actions that we execute on them. Two entities are included in the same category, that is, they evoke the same pattern of neural activity, if we tend to execute the same actions on them, whereas two entities are included in different categories if we tend to execute different actions on them.

If we interpret not only signs but also their meaning in terms of motor actions, we can understand how meanings can be shared between speakers and hearers. Motor actions are the only aspect of behavior which is inter-individually accessible. A has no direct access to what B perceives, feels, or thinks, but only to how B moves its body and to the consequences of these movements. Meanings can be shared if the categories they refer to are represented in terms of motor actions in the brain of both speaker and hearer. Mirror neurons can play a role not only with respect to the motor (phono-articulatory) actions that result in the production of vocal signs but also with respect to the motor actions of all kinds that we execute on the entities that vocal signs refer to. A gestural origin of human language may have facilitated the emergence of shared meanings. As Arbib recognizes, gestural signs are more iconic than vocal signs, which means that gestural signs are motor actions which physically resemble the motor actions that we execute on the entities they refer to. Vocal signs are arbitrary, that is, non-iconic, but they may have exploited the already existing shared meanings neurally represented in terms of inter-interindividually accessible motor actions executed on objects.

Artificial life simulations and experiments suggest that seeing objects or processing words referring to objects automatically activates canonical actions we perform on them, particularly reaching and grasping movements. Borghi and colleagues (Borghi et al. 2002, 2005, submitted; Di Ferdinando & Parisi 2004) evolved simulated organisms using a genetic algorithm (Holland 1992). Each organism lives in a bidimensional environment containing four objects, either upright or reversed, with a handle protruding on the right or on the left. The organism possesses a visual system allowing it to see different objects, one at a time, and a motor system consisting of a single arm composed of two movable segments; the arm sends proprioceptive information to the organism, specifying the arm's current position. The organism's nervous system was simulated with a neural network (Fig. 1).

Organisms learned to reach the handle of the object independently of its position (Task 1) and then they learned to reach one of two buttons located below the handle to decide whether the object was upright or reversed (Task 2). In one condition, the button to be reached was on the same side of the object's handle; in another condition, it was on the opposite side. Task 1 reproduced real-life experience; Task 2 replicated an experiment made by Tucker and Ellis (1998). When the handle location and the button were spatially compatible, learning occurred earlier (in terms of generations required to reach an optimal performance) than when they were not. The results suggest that affordances of objects become learned in association with successful reaching movements. Once reaching becomes established, seeing the handle of objects activates appropriate movements.

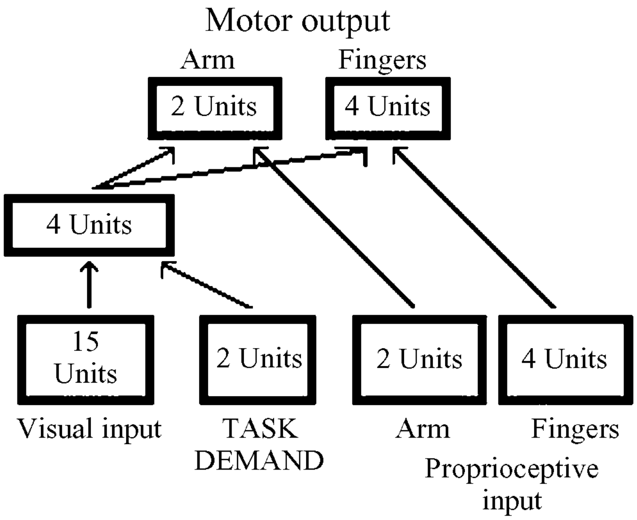


Figure 1 (Parisi et al.). Neural network controlling an arm.

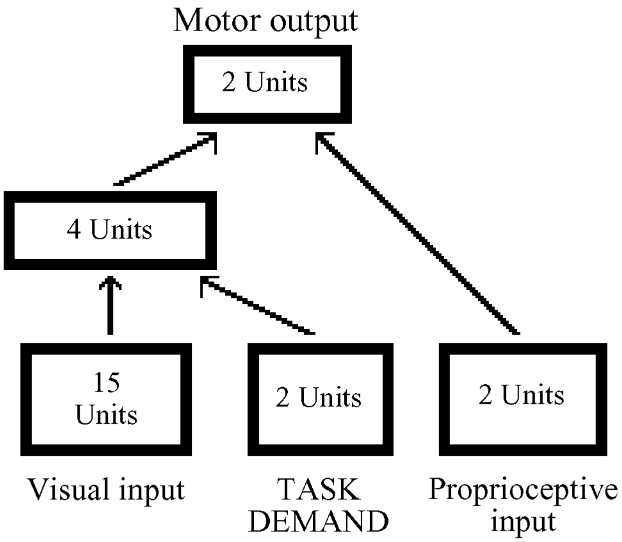


Figure 2 (Parisi et al.). Neural network controlling an arm and a hand.

Tsiotas et al. (in press) simulated an organism with a nervous system (Fig. 2) and with an arm terminating with a hand composed of two fingers, a thumb and an index, each composed by two segments.

The organism lived in a bidimensional environment containing four objects, either large or small and red or blue. First, the organism had to learn to grasp small objects with a precision grip and large objects with a power grip, then to decide the objects' color by grasping a small or a large button. Learning occurred earlier when the grip required to respond to the object and to decide the color was the same than when it was not, even if object size was irrelevant for the task (Ellis & Tucker 2000).

The inter-accessibility of these simple gestures which are automatically activated by objects, may have played a relevant role for language evolution. Crucially, these gestures are automatically activated not only by visual stimuli but by words, too (Gentilucci 2003b; Tucker & Ellis 2004). Borghi et al. (2004) found in a part-verification task that responding by moving the arm in a direction incompatible with the part location (e.g., responding downward to verify that a car has a roof, upward to verify that a car has wheels) was slow relative to responding in a direction compatible with the part location.

The presence of action-based compatibility effects also with words, argues for the involvement not only of the dorsal but also of the ventral system and of long-term knowledge in generating affordances: accordingly, these effects would be accounted for by long-term visuomotor associations between objects and actions executed on them.

An avian parallel to primate mirror neurons and language evolution?

Irene M. Pepperberg

Department of Psychology, Brandeis University, MS062, Waltham, MA 02454. **impepper@media.mit.edu** **http://www.alexfoundation.org**

Abstract: Arbib presents a reasoned explanation for language evolution from nonhuman to human primates, one that I argue can be equally applied to animals trained in forms of interspecies communication. I apply his criteria for language readiness and language (in actuality, protolanguage) to the behavior of a Grey parrot (*Psittacus erithacus*) taught to communicate with humans using rudiments of English speech.

Arbib approaches an old chestnut – language evolution – from the novel standpoint of mirror neurons (MN), building upon his earlier theses. His counterarguments for innatist theories are clearly on target. With little to critique, I focus on possible parallels between Arbib's proposals and Grey parrot behavior – particularly that of my oldest subject, Alex (Pepperberg 1999).

Concerning Arbib's criteria for language-readiness (LR), little is unique to primates. Arbib provides not LR but CCR – "complex communication-ready" – criteria. He suggests this possibility but omits details. LR1 (complex imitation), reproduction of novel behavior that can be approximated by existent actions and their variants, is demonstrated, for example, by Alex's initial immediate utterance of "s(pause)wool" for "spool" (for a wooden bobbin; Pepperberg 2005b, /p/ being particularly difficult to produce without lips and teeth (Patterson & Pepperberg 1998). LR2 (symbolization), LR3 (parity), and LR4 (intention) are demonstrated in detailed studies of Alex's referential communication (Pepperberg 1999). LR5 (temporal versus hierarchical ordering) is more difficult to prove, except possibly in the understanding and use of interactive dialogue (Pepperberg 1999). LR6 (past/future) occurs in any animal that can be operantly conditioned. Although few data exist on Grey parrot behavior in nature, LR7 is likely, given that Greys take several years to reach sexual maturity.

In LA1 through LA4, Arbib also focuses on primates, but Greys seemingly meet most criteria. For LA1, for example, Alex transfers the use of "none" from responding to "What's same/different?" for two objects when nothing is same or different, to responding to, without training, "What color bigger?" for equally sized objects (Pepperberg & Brezinsky 1991), and then, again without training, to designate an absent quantity in an enumeration task (Pepperberg & Gordon 2005). Furthermore, to Alex, "paper," for example, is not merely index card pieces used for initial training, but large sheets of computer output, newspapers, and students' textbooks. For LA2, Alex comprehends recursive, conjunctive queries (e.g., "What object is green and 3-corner?" versus "What color is wood and 4-corner?" versus "What shape is blue and wool?"; Pepperberg 1992). LA3 has not been demonstrated directly in Greys, but birds likely have episodic memory (e.g., work by Clayton et al. 2003). LA4, learnability, exists with respect to semantics and, to a limited extent, for sentence frames (appropriate use of "I want X" versus "Wanna go Y"; Pepperberg 1999). Interestingly, Arbib's criteria closely parallel Hockett's (1959) design features; direct comparison would be instructive.

Given these parallels, do Grey parrots also have MN systems – neurons that, for example, react similarly when birds hear *and* speak human labels? Biologically, existent evidence is sparse but intriguing. For oscine birds' own song, some parallels exist with primates. Songbirds' high vocal center (HVC) sends efferents to both input and output branches of the song system; HVC is necessary for song production and has neurons showing song-specific auditory responses (Williams 1989). Furthermore, playback of birds' own song during sleep causes neural activity comparable to actual singing (Dave & Margoliash 2000).

How these findings relate to parrot brains, which are organized differently from songbird brains (e.g., Jarvis & Mello 2000; Striedter 1994) is unclear. Although studies of ZENK gene[1] expression show separation of budgerigar (*Melopsittacus undulates*) response regions for hearing and vocalizing warble song (Jarvis & Mello 2000), electrophysiological studies in the frontal neostriatum of awake budgerigars show activity both in production of and response to calls (Plumer & Striedter 1997; 2000); evidence also exists for additional budgerigar auditory-vocal pathways (e.g., Brauth et al. 2001). Because ZENK response apparently is tuned to specific song features (Ribeiro et al. 1998), the relevance of these data for MNs in talking parrots is unknown.

However, arguments for complex imitation, and by inference, brain structures to support such behavior, exist. Like children described by Arbib, Alex goes beyond simple imitation; he acquires the phonological repertoire, some words, and basic "assembly skills" of his trainers and appears to parse complex behavior patterns (words and phrases) into recombinable pieces and familiar (or semi-familiar) actions. In addition to material described above, Alex (1) recognizes and produces small phonetic differences ("tea" vs. "pea") meaningfully (Patterson & Pepperberg 1994; 1998), (2) produces initial phonemes differently depending upon subsequent ones (/k/ in "key" vs. "cork"; Patterson & Pepperberg 1998), and (3) consistently recombines parts of labels according to their order in existent labels (i.e., combines beginnings of one label with the ends of others – e.g., "banerry" [for apples] from banana/cherry. After analyzing more than 22,000 vocalizations, we never observed backwards combinations such as "percup" instead of "cupper/copper"; Pepperberg et al. 1991).

Surprisingly, Arbib doesn't discuss Greenfield's (1991) studies that might also involve co-opting gestural forms for vocal language, although she does not examine MNs and imitation. Apparently, human children – and language-trained chimps, but not monkeys – simultaneously develop hierarchical object and linguistic ordering (e.g., serial cup stacking, phrases like I + want + X) as, Greenfield argues, a consequence of Broca/F5 maturation. MNs in these brain areas are activated by both action and observation of hand or mouth gestures; less advanced MNs exist in monkeys than in apes and humans. Similar behavior is observed in Grey parrots (Pepperberg & Shive 2001), although avian combinations both involve the beak. Greenfield implies that these actions emerge without overt instruction; however, these behavior patterns are likely observed from birth (or hatching). Maybe only after maturation of MN and canonical neuron systems can they be expressed (Pepperberg 2005a).

In sum, the communication system I have taught Grey parrots will never be fully congruent with any current human language, but I am intrigued by the many parallels that can be drawn between their protolanguage and that described by Arbib for early *Homo*: Start with a brain of a certain complexity and give it enough social and ecological support; that brain will develop at least the building blocks of a complex communication system.

ACKNOWLEDGMENT

Preparation of this commentary was supported by donors to The Alex Foundation.

NOTE

1. Expression of the ZENK gene, a songbird analog to a human transcription factor, egr-1, is driven by actions of singing and hearing. Hence, it is used to form a functional map of avian brains for behavior related to both auditory processing and vocal production (Jarvis & Mello 2000).

Contagious yawning and laughing: Everyday imitation- and mirror-like behavior

Robert R. Provine

Department of Psychology, University of Maryland Baltimore County, Baltimore, MD 21250. **provine@umbc.edu**

Abstract: Infectious yawning and laughing offer a convenient, noninvasive approach to the evolution, development, production, and control of imitation-like and mirror-like phenomena in normal, behaving humans.

The analysis of a scientific problem can benefit from taking a broad perspective before turning to narrower and more reductive issues. In this spirit, I nominate contagious yawning and laughing for consideration, which are two of the most familiar cases of human behavior with imitation-like and mirror-like properties. Even their relegation to special-case status would help set parameters and inform readers who are more familiar with these acts than such esoteric and inaccessible phenomena as mirror neurons. An attractive feature of contagious yawning and laughing as scientific problems is that we can use ourselves as subjects – no electrophysiological laboratory is required. They also offer tantalizing insights into the evolutionary process through which a motor act may become mirrored or imitated.

Contagious yawning and laughing involve a chain reaction of behavior and physiology that propagates through and synchronizes the state of a group. Being unconsciously controlled, the contagious responses do not involve a desire to replicate an observed yawn or laugh – we just do them. Although the sensory vector for contagious yawns is primarily visual and that for laughter is primarily auditory, both contagious acts involve the replication of observed movements, whether the facial contortions of the yawn, or the respiratory movements that produce the vocalization of laughter.

Although the focus of this commentary is on the mirror-like and imitation-like properties of contagion, the analysis of mechanism must begin with the motor act brought under stimulus control. Yawns and laughs evolved before the stimulus triggers responsible for their contagion. This is a case of *motor precocity*, the common tendency of motor systems to develop or evolve prior to receiving sensory inputs. Organisms often "spond before they respond." Motor systems can be adaptive, stand-alone processes, unlike sensory systems that, by themselves, lack adaptive significance because they have no behavioral consequence. (By extension, reflexes are unlikely to emerge de novo because they require the improbable simultaneous genesis of both a sensory and motor process.) Let us now consider the evolution of yawning and laughing and how they came under sensory control.

Yawning (Provine 1986) is an ancient, stereotyped motor pattern that is performed by most vertebrates and develops prenatally in humans. Once initiated, a yawn goes to completion – recall the difficulty of stifling a yawn. There are no half-yawns. The motor pattern generator for yawning probably resides in the brain stem along with other pulmonary and vasomotor control centers. A yawn, like a laugh, is not under voluntary control and cannot be produced on command.

Contagious yawning (Provine 1986; 1989) probably emerged many millions of years after the ubiquitous motor act and, although it may be present in other species, has been clearly demonstrated only in humans. Lacking the remarkable precocity of the motor act, contagious yawning of humans appears sometime during early childhood, a developmental trajectory that suggests the involvement of a separate and higher brain mechanism. Contagious yawns can be triggered by the observation of the overall configuration of the animate, yawning face, regardless of its axial orientation or presence of the gaping mouth. (Shielding a yawn will not block its contagion.) The neurological yawn detector is so broadly tuned that almost any stimulus associated with yawning can trigger the act, including, as some readers have noticed, even thinking about or reading about yawning. The broad tuning insures contagion in darkness or in the absence of line-of-sight visual contact with a yawner.

Laughter has a clearer and much shorter history than yawning and is associated with the evolution of social play in mammals (Provine 1996; 2000). Laughter is literally the sound of labored breathing in rough and tumble play, where the sound of panting has come to represent the playful act that produced it. Ethologists refer to such processes as ritualization. Laughter evolved as a play vocalization, an unconsciously controlled, therefore *honest signal* that an encounter has playful intent and is not a physical assault. In humans, the "pant-pant" laughter of our primate ancestors morphed into "ha-ha." Laughter is the clearest example of how a vocalization evolved – it does not involve the arbitrary pairing of a sound with a meaning. (The transition from "pant-pant" to "ha-ha" laughter reflects the increased vocal control of humans enabled by bipedality and ultimately explains why we can speak and other great apes cannot.) Laughter and speech time-share the same vocal apparatus, but each maintains unique features and neurological mechanisms. Laughter lacks the voluntary control of spoken words, and we tend to either laugh or speak, with speech being dominant because laughter seldom interrupts the phrase structure of speech. Laughter *punctuates* the speech stream (Provine 1993).

Laughter triggers the laughter of those who hear it, synchronizing and amplifying the neurobehavioral status of a group. It is the basis of the notorious television laugh tracks. Crying is another infectious vocalization, at least among human infants (Simner 1971). As suggested by Arbib, such processes are probably common among animals. Contagious laughs occur almost immediately after the stimulus laugh, in contrast to contagious yawns where there is a gradual increase in the probability of yawning during the seconds after the observed yawn.

A challenge of comparing the mirror systems of Arbib with those of yawning and laughter is that so little is known about the neurology of the latter. The laughing/yawning systems may, for example, more resemble systems involved in monkey vocalizations (midbrain and cingulate cortex) than those for language (e.g., Broca's and Wernicke's regions) or the specific mirror system considered by Arbib, the hand and orofacial system of monkey premotor area F5. However, the yawning/laughter systems may be a convenient exemplar of a class of processes at the foundation of Arbib's proposal that can teach us about mirror/imitation mechanisms and their evolution. The parsimony of biological systems suggests that, in whole or in part, standard processes, components, and circuits in the neurological tool kit are likely to find many applications.

Motivation rather than imitation determined the appearance of language

Pavel N. Prudkov

Ecomon Ltd., Selskohosyastvennaya ul. 12-a, Moscow, Russia. **pnprudkov@mtu-net.ru**

Abstract: Arbib derives the origin of language from the emergence of a complex imitation system; however, it is unlikely that this complication could occur without a prior complicating within the imitated systems. This means that Arbib's hypothesis is not correct, because the other systems determined the appearance of language. In my opinion, language emerged when the motivational system became able to support goal-directed processes with no innate basis.

In the target article Arbib derives the origin of language from the emergence of a complex imitation system among ancient *Homo*. Describing in detail how the complex imitation system could facilitate the formation of protosign and protospeech, he says nothing, however, about why this system must have emerged. This is a serious problem; imitation is, by definition, copying of other processes, therefore the complexity of the imitation system of an organism cannot exceed the complexity of the systems to be imitated. This principle seriously constrains the possibility of the emergence of a new, more complex imitation system without the corresponding complicating within the systems to be imitated. Such a possibility seems to underlie Arbib's approach because, in emphasizing the changes in the imitation system, he does not require similar fundamental changes in other systems.

Of course, it is impossible to abandon the idea that the complex imitation system could emerge as a result of a single mutation without the corresponding changes in other systems of some ancient hominids; but such hominids occasionally benefited from their new possibilities, thereby surviving successfully, until other systems achieved the complexity of the imitation system; and then natural selection started working more conventionally again. The probability of this scenario is extremely low, obviously. Another approach to the origin of the complex imitation system, which seems much more probable, is that a certain complication of other systems preceded this system and made its appearance necessary. This, however, means that Arbib's hypothesis suggesting that the complex imitation system is the "missing link" is not correct, because other systems in fact determined the appearance of language.

Like other hypotheses of language origin, Arbib's hypothesis is based on the idea that language is a means of communication. This definition is correct but incomplete: language is a means of communication for people engaged in a joint activity. There is a clear correlation between the diversity of activities and the complexity of the language serving these activities. Modern languages consist of hundreds of thousands of words only because these languages are applied in thousands of diverse activities. Each human activity is goal-directed, hence, the complexity of languages is a consequence of the ability of the human brain to construct diverse goals. Indeed, most human goals are not constrained by any innate basis; they are social, and result from interactions between people. So, there is an obvious connection between language and the ability to construct and maintain long-term motivations with no innate basis.

No nonhuman animals have a motivational system with similar characteristics. Animals have long-term motivations (e.g., sex, hunger), but these are all innate. An animal can form learned motivations, but only when its basic drives are activated. The hypothesis that the motivation of animals is always constrained by the activation of basic drives was suggested by Kohler (1917/1927), and despite intensive researches, there have still been no data inconsistent with it (Suddendorf & Corballis 1997). With the limited and stable number of long-term motivations, animals are constrained in using and developing their languages. Since all their motivations are connected with vital functions, any serious misunderstanding in the process of communication can be fatal; as a result, the number of signals in animal languages must be limited, and the signals must have unequivocal meanings. Roughly speaking, animals do not have a language similar to human languages because they simply do not need it.

I have suggested elsewhere that the emergence of the ability to construct and maintain long-term goals with no innate basis was the missing link for language (Prudkov 1999c) and for the other distinctively human characteristics (Prudkov 1999a; 1999b) because the ability allowed ancient humans to overcome the constraints of innate motivations, thus providing the possibility of constructing new, flexible, and open systems. In other words, protolanguage emerged because in new situations conditioned by goals having no innate basis, the innate communicative means became inefficient for interactions between ancient hominids, and those who were able to construct new means succeeded in reproduction. Of course, language, imitation, and the theory of mind had started evolving then. It is very important to emphasize that without the prior (or parallel) formation of the system able to construct learned, long-term motivations, any changes in other systems (e.g., in intelligence) were not sufficient to overcome innate constraints. For example, the capacity of birds to navigate in three-dimensional space on the basis of visual cues obviously exceeds that of humans, but innate mechanisms determine the behavior of birds.

It is reasonable to think that there was a reciprocal interaction in the evolution of human language and motivation. The new motivational ability spurred the development of language; afterwards language was used to construct efficient, purposeful processes, and this interaction likely determined all stages of human evolution. This joint evolution was facilitated by the fact that a common mechanism that evolved within these systems is the capacity to form and execute complex, hierarchical, goal-directed processes (such processes are rapid and relatively simple in language and are slow and complex in motivation) (Prudkov & Rodina 1999). In other words, I agree with Arbib that humans have a language-ready brain rather than special mechanisms embedded in the genome. The capacity was also involved in the development of the imitation system, because a basic characteristic distinguishing the human imitation system from its animal analogs is the possibility to imitate more complex and long-term processes. But the development of the imitation system itself is not sufficient to construct protolanguage, because only the new motivational system could make imitation voluntary and arbitrary. Indeed, in emphasizing that at a certain stage of evolution communication became voluntary and intentional, Arbib does not explain what mechanisms underlay such possibilities of communication.

In my opinion, the gestural and vocal components of protolanguage emerged together, but the latter gained advantage in the development because, unlike gestures, which are effective only in dyadic contacts, vocalizations are more effective in group actions (group hunting, collective self-defense, etc.), which became the first actions guided by goals having no innate basis.

Vocal gestures and auditory objects

Josef P. Rauschecker

Laboratory of Integrative Neuroscience and Cognition, Georgetown University School of Medicine, Washington, DC 20057-1460.
rauschej@georgetown.edu

Abstract: Recent studies in human and nonhuman primates demonstrate that auditory objects, including speech sounds, are identified in anterior superior temporal cortex projecting directly to inferior frontal regions and not along a posterior pathway, as classically assumed. By contrast, the role of posterior temporal regions in speech and language remains largely unexplained, although a concept of vocal gestures may be helpful.

In his target article, Arbib maintains (and before him, Rizzolatti & Arbib 1998) that language originated from a system of mirror neurons coding manual gestures, rather than from vocal communication systems present in nonhuman primates (and other animals). I do not doubt the usefulness of the mirror-neuron concept, which brings back to mind the motor theory of speech perception (Liberman et al. 1967). In fact, many recent neuroimaging studies have independently demonstrated a simultaneous activation of what were previously thought of as separate centers for the production and perception of human language, Broca's and Wernicke's areas, respectively. These designations go back more than a century to crudely characterized single-case studies of neurological patients, which have been shown by modern magnetic resonance imaging (MRI) techniques (Bookheimer 2002) to have missed much more brain than the relatively small regions that now bear their discoverers' names.

Both on that basis and on the basis of his own belief in intertwined systems of perception and action, it is surprising that Arbib continues to use this outdated terminology. "Broca's area" at least is redefined by him as part of a system that deals with, among others, "sequential operations that may underlie the ability to form words out of dissociable elements" (sect. 8), a definition that many researchers could agree with, although the exact correspondence with cytoarchitectonically defined areas and the homologies between human and nonhuman primates are still controversial. "Wernicke's area," by contrast, gets short shrift. Arbib talks about it as consisting of the posterior part of Brodmann's area 22, including area Tpt of Galaburda and Sanides (1980) and an "extended [parietal area] PF," suggesting that this is the only route that auditory input takes after it reaches primary auditory cortex. Of course, this suggestion echoes the classical textbook view of a posterior language pathway leading from Wernicke's to Broca's area via the arcuate fascicle.

A remarkable convergence of recent neurophysiological and functional imaging work has demonstrated, however, that the analysis of complex auditory patterns and their eventual identification as auditory objects occurs in a completely different part of the superior temporal cortex, namely, its anterior portion. The anterior superior temporal (aST) region, including the anterior superior temporal gyrus (STG) and to some extent the dorsal aspect of the superior temporal sulcus (STS), project to the inferior frontal (IF) region and other parts of the ventrolateral prefrontal cortex (VLPFC) via the uncinate fascicle. Together, the aST and IF cortices seem to form a "what" stream for the recognition of auditory objects (Rauschecker 1998; Rauschecker & Tian 2000), quite similar to the ventral stream for visual object identification postulated previously (Ungerleider & Mishkin 1982). Neurophysiological data from rhesus monkeys suggest that neurons in the aST are more selective for species-specific vocalizations than are neurons in the posterior STG (Tian et al. 2001). In humans, there is direct evidence from functional imaging work that intelligible speech as well as other complex sound objects are decoded in the aST (Binder et al. 2004; Scott et al. 2000; Zatorre et al. 2004).

It seems, therefore, that the same anatomical substrate supports both the decoding of vocalizations in nonhuman primates and the decoding of human speech. If this is the case, the conclusion is hard to escape that the aST in nonhuman primates is a precursor of the same region in humans and (what Arbib may be reluctant to accept) that nonhuman primate vocalizations are an evolutionary precursor to human speech sounds. Indeed, the same phonological building blocks (or "features"), such as frequency-modulated (FM) sweeps, band-passed noise bursts, and so on, are contained in monkey calls as well as human speech. Admittedly, the decoding of complex acoustic sound structure alone is far from sufficient for language comprehension, but it is a necessary precondition for the effective use of spoken speech as a medium of communication. Arbib argues, with some justification, that communication is not bound to an acoustic (spoken) medium and can also function on the basis of visual gestures. However, in most hearing humans the acoustic medium, that is, "vocal gestures," have gained greatest importance as effective and reliable carriers of information.

An interesting question remaining, in my mind, is, therefore, how the auditory feature or object system in the aST could interact with a possible mirror system, as postulated by Arbib and colleagues. The projection from aST to IF seems like a possible candidate to enable such an interaction. Indeed, auditory neurons, some of them selectively responsive to species-specific vocalizations, are found in the VLPFC (Romanski & Goldman-Rakic 2002). According to our view, aST serves a similar role in the auditory system as inferotemporal (IT) cortex does for the visual system. Which role, if any, Wernicke's area (or posterior STG) plays for vocal communication, including speech and language, remains the bigger puzzle. Understanding it as an input stage to parietal cortex in an auditory dorsal pathway is a good hint. However, as Arbib would say, "empirical data are sadly lacking" and need to be collected urgently.

Continuities in vocal communication argue against a gestural origin of language

Robert M. Seyfarth

Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104. **seyfarth@psych.upenn.edu**
http://www.psych.upenn.edu/~seyfarth/Baboon%20research/index.htm

Abstract: To conclude that language evolved from vocalizations, through gestures, then back to vocalizations again, one must first reject the simpler hypothesis that language evolved from prelinguistic vocalizations. There is no reason to do so. Many studies – not cited by Arbib – document continuities in behavior, perception, cognition, and neurophysiology between human speech and primate vocal communication.

Arbib argues that the emergence of human speech "owes little to nonhuman vocalizations" and concludes that "evolution did not proceed directly from monkey-like primate vocalizations to speech but rather proceeded from vocalization to manual gesture and back to vocalization again" (target article, sect. 2.3). Accepting this hypothesis requires us to adopt a convoluted argument over a simple one. There is no need to do so.

If dozens of scientists had been studying the natural vocalizations of nonhuman primates for the past 25 years and all had concluded that the vocal communication of monkeys and apes exhibited no parallels whatsoever with spoken language, one might be forced to entertain Arbib's hypothesis. If years of neurobiological research on the mechanisms that underlie the perception of calls by nonhuman primates had revealed no parallels with human speech perception, this, too, might compel us to reject the idea that human language evolved from nonhuman primate vocalizations. Neither of these conclusions, however, is correct.

Arbib offers his hypothesis as if he had carefully reviewed the literature on nonhuman primate vocal communication and thoughtfully rejected its relevance to the evolution of human language. Readers should be warned, however, that his review ends around 1980 and even neglects some important papers published before that date.

Primate vocal repertoires contain several different call types that grade acoustically into one another. Despite this inter-gradation, primates produce and perceive their calls as, roughly speaking, discretely different signals. Different call types are given in different social contexts (e.g., Cheney & Seyfarth 1982; Fischer 1998; Fischer et al. 2001a; Hauser 1998; Snowdon et al. 1986). In playback experiments, listeners respond in distinct ways to these different call types, as if each type conveys different information (e.g., Fischer 1998; Fischer et al. 2001b; Rendall et al. 1999). Listeners discriminate between similar call types in a manner that parallels – but does not exactly duplicate – the categorical perception found in human speech (Fischer & Hammerschmidt 2001; Owren et al. 1992; Prell et al. 2002; Snowdon 1990; Zoloth et al. 1979). Offering further evidence for parallels with human speech, the grunts used by baboons (and probably many other primates) differ according to the placement of vowel-like formants (Owren et al. 1997; Rendall 2003).

Arbib incorrectly characterizes primate vocalizations as "involuntary" signals. To the contrary, ample evidence shows that nonhuman primate call production can be brought under operant control (Peirce 1985) and that individuals use calls selectively in the presence of others with whom they have different social relations (for further review and discussion, see Cheney & Seyfarth 1990; Seyfarth & Cheney 2003b).

Because nonhuman primates use predictably different calls in different social and ecological contexts, listeners can extract highly specific information from them, even in the absence of any supporting contextual cues. For example, listeners respond to acoustically different alarm calls as if they signal the presence of different predators (Fichtel & Hammerschmidt 2002; Fischer 1998; Seyfarth et al. 1980), and to acoustically different grunts as if they signal the occurrence of different social events (Cheney & Seyfarth 1982; Rendall et al. 1999). In habituation-dishabituation experiments that asked listeners to make a same-different judgment between calls, subjects assessed calls based on their meaning, not just their acoustic properties (Cheney & Seyfarth 1988; Zuberbuhler et al. 1999). The parallels with children's perception of words cannot be ignored (see Zuberbuhler 2003 for review).

Indeed, it is now clear that although primates' production of vocalizations is highly constrained, their ability to extract complex information from sounds is not (Seyfarth & Cheney 2003b). Upon hearing a sequence of vocalizations, for example, listeners acquire information that is referential, discretely coded, hierarchically structured, rule-governed, and propositional (Bergman et al. 2003; Cheney & Seyfarth, in press). These properties of primates' social knowledge, although by no means fully human, bear striking resemblances to the meanings we express in language, which are built up by combining discrete-valued entities in a structured, hierarchical, rule-governed, and open-ended manner. Results suggest that the internal representations of language meaning in the human brain initially emerged from our prelinguistic ancestors' knowledge of social relations, as exhibited in the information they acquire from vocalizations (Cheney & Seyfarth 1997; in press; Worden 1998).

Nonhuman primate vocalizations also exhibit parallels with human speech in their underlying neural mechanisms. Behavioral studies of macaques suggest that the left hemisphere is specialized for processing species-specific vocalizations but not other auditory stimuli (Hauser & Anderson 1994; Petersen et al. 1978). Lesion results demonstrate that ablation of auditory cortex on the left but not the right hemisphere disrupts individuals' ability to discriminate among acoustically similar call types (Heffner & Heffner 1984). Most recently, Poremba et al. (2004) measured local cerebral metabolic activity as macaques listened to a variety of auditory stimuli. They found significantly greater activity in the left superior temporal gyrus as compared with the right, but only in response to conspecific vocalizations. These and other results (e.g., Wang et al. 1995; see Hauser [1996] and Ghazanfar & Hauser [2001] for review) suggest that Arbib is wrong to assume that primate vocalizations "appear to be related to non-cortical regions" (sect. 1. 2, para. 3). They further suggest that the neurophysiological mechanisms underlying human speech processing evolved from similar mechanisms in our nonhuman primate ancestors.

In sum, research demonstrates a striking number of continuities – in behavior, perception, cognition, and neurophysiology – between human speech and the vocal communication of nonhuman primates. Nonhuman primate vocal communication does not qualify as language, but it does exhibit many of the characteristics that one would expect to find if human language had evolved from the vocal communication and cognition of the common ancestor of human and nonhuman primates.

Arbib cites none of this research. As a result, his presentation is strongly biased in favor of his own view that the emergence of human speech "owes little to nonhuman vocalizations" (target article, Abstract). To accept the convoluted hypothesis that spoken language evolved from vocalizations, through gestures, then back to vocalizations again, one must first have good reason to reject the simpler hypothesis that spoken language evolved from prelinguistic vocal communication. A substantial body of data argues against such a rejection.

Making a case for mirror-neuron system involvement in language development: What about autism and blindness?

Hugo Théoret[a] and Shirley Fecteau[b]

[a]*Departement de Psychologie, Université de Montréal, Centre Ville, Montreal, Qc H3C 3J7, Canada;* [b]*Faculté de Médecine, Université de Montréal, Centre Ville, Montreal, Qc, H3C 3J7, Canada.* **hugo.theoret@umontreal.ca shirley.fecteau@umontreal.ca**

Abstract: The notion that manual gestures played an important role in the evolution of human language was strengthened by the discovery of mirror neurons in monkey area F5, the proposed homologue of human Broca's area. This idea is central to the thesis developed by Arbib, and lending further support to a link between motor resonance mechanisms and language/communication development is the case of autism and congenital blindness. We provide an account of how these conditions may relate to the aforementioned theory.

Arbib presents a strong argument in favor of a link between mirror neurons (MN), imitation, and the development of human language. We endorse his thesis that a protolanguage based on manual gestures was a precursor to human language as we know it today. Additional support for this claim comes from two seemingly different conditions: autism and congenital blindness.

Autism. Language and communication deficits are one of the defining features of autism spectrum disorders (ASD; American Psychiatric Association 1994) and are core elements of their diagnosis and prognosis (Herbert et al. 2002; Ventner et al. 1992). Particularly relevant is the fact that these impairments are more prominent in pragmatic speech associated with social communication (Tager-Flusberg 1997). Interestingly, individuals with ASD also display well-documented deficits in imitative behavior (e.g., Avikainen et al. 2003). Recent magnetoencephalographic data suggest that an abnormal mirror-neuron system (MNS) may underlie the imitative impairment observed in individuals with ASD (Nishitani et al. 2004). That study reported imitation-related abnormalities in Broca's area and its contralateral homologue, the human equivalent of monkey area F5, where most MN are found (Rizzolatti & Craighero 2004).

The idea that imitative abilities, and possibly language impairments, are related to basic MN dysfunction in ASD was recently investigated in our laboratory. In line with the MN hypothesis of ASD (Williams et al. 2001), motor cortex activation during the observation of simple finger movements was found to be significantly weaker in individuals with ASD compared to matched controls (Théoret et al. 2005). The MNS/language disorder hypothesis is also supported by the fact that individuals with autism display structural abnormalities in Broca's area (Herbert et al. 2002). Other symptoms that may be associated in some way with MN dysfunction in ASD include abnormal eye gaze, theory-of-mind deficits, the use of other's hands to communicate or demand, hand mannerisms, repetitive behaviors, and echolalia.

Taken together, these data support Arbib's main argument that a simple action observation/execution matching mechanism anchored in area F5 (Broca's area in humans) may have evolved into a complex system subserving human language. Consequently, a pathological, congenital dysfunction of the mirror-cell system in humans would be expected to dramatically affect social interactions and language/communication as a result of gesture/speech interpretation and acquisition. This appears to be the case in ASD. As mentioned by Arbib, Broca's is not the only area making up the human language and MNS. It is thus possible that other regions within the MNS underlie some intact language skills in some ASD individuals (e.g., grammar and syntax), which could in turn partly account for the heterogeneity of the symptoms across individuals.

To that effect, the case of individuals with ASD and normal IQ is particularly relevant to the argument put forth by Arbib. In that population, it is the social and pragmatic aspects of language that are usually impaired, with some individuals displaying normal abilities in, for example, vocabulary and syntax. It appears that abnormal imitation-related cortical activations in ASD with normal IQ are located mostly within the inferior frontal gyrus as opposed to the superior temporal sulcus and inferior parietal lobule (Nishitani et al. 2004). In light of these neurophysiological results, it may be that modularity of the MNS can account for differential language symptomatology, with Broca's area being the principal component of the system.

Blindness. Another pathological condition that may add some insight into the perspective offered by Arbib is congenital blindness. It has been suggested that congenitally blind individuals display autism-like characteristics (Hobson & Bishop 2003). For example, visually impaired children perform at lower levels than normal subjects on theory-of-mind tasks (Minter et al. 1998), and blind children are at an increased risk of meeting diagnostic criteria for autism (Brown et al. 1997). Interpretation of these data as suggesting a causal link between sensory deprivation and autism-like characteristics has been challenged (Baron-Cohen 2002), but they nevertheless bring to mind interesting questions regarding ASD, MN function, and language impairment.

Some blind children display fewer periods in which they direct language towards other children and are generally impaired in the social and pragmatic aspects of language (Hobson & Bishop 2003), reminiscent of individuals with ASD. In blind individuals, lack of visual input would derail the normal mechanism matching action perception and execution within the visual system. A motor resonance mechanism could still operate through the auditory modality (Kohler et al. 2002), but in an obviously limited manner due to lack of visual input.

Mechanisms of disorder. We have tried to describe two pathological conditions that offer insight into the role of the MNS in language/communication. We have showed that a breakdown in MN function may be associated with specific language impairments, most notably pragmatic speech. In contrast to the theory put forth by Arbib, these examples speak to the ontogeny, rather than the phylogeny, of language. Nevertheless, they share a striking similarity: the necessity of an adequately "evolved" (as Arbib puts it) MNS to develop the unique ability of human language. Although still speculative, the two conditions we have described suggest different mechanisms that may lead to MNS impairment and associated language deficits.

In the case of blindness, it may be that loss of visual input impairs the normal development of a motor resonance system, thereby leading to language/communication deficits. In that sense, it is an environmental factor that hinders adequate development of the MNS. In ASD, where genetic factors are an important part of the etiology, individuals may be born with a dysfunctional MNS, preventing normal language and social behavior. In that regard, it is tempting to look at the Forkhead box P2 (FOXP2) gene, located on chromosome 7q, which is believed to be implicated in the acquisition of language (Lai et al. 2001) and may be involved in the human properties of the MNS (Corballis 2004). Most evidence argues against a direct link between autism and FOXP2 (e.g., Newbury et al. 2002), but the idea that MN development may be genetically determined is an intriguing possibility that requires further investigation.

In summary, this commentary highlights the need to test Arbib's theory against various pathological conditions, either those specific to language (e.g., aphasia) or those which may be associated with MN dysfunction (autism, schizophrenia, Williams' syndrome). For example, one of the co-morbidities of specific language impairment (SLI) is motor impairment (Hill 2001), suggesting yet another association between motor skill and language dysfunction. It seems obvious to us that specific predictions of Arbib's model need to be tested this way, as direct evidence in support of some aspects of the theory is lacking.

ACKNOWLEDGMENTS

This commentary was supported by grants from the National Sciences and Engineering Research Council of Canada and the Fonds de Recherche en Santé du Québec.

Language is fundamentally a social affair

Justin H. G. Williams

Department of Child Health, University of Aberdeen School of Medicine, Royal Aberdeen Children's Hospital, Aberdeen AB25 2ZG, Scotland, United Kingdom. **justin.Williams@abdn.ac.uk**
http://www.abdn.ac.uk/child_health/williams.hti

Abstract: Perhaps the greatest evolutionary advantage conferred by spoken language was its ability to communicate mentalistic concepts, rather than just extending the vocabulary of action already served by an imitation function. An appreciation that the mirror-neuron system served a simple mentalising function before gestural communication sets Arbib's theory in a more appropriate social cognitive context.

It may not be an obvious question to ask why spoken language should evolve from gestural communication, but it is an important one. Simply put, if gesture can be used to communicate effectively, why evolve speech? Why didn't we just evolve a complex gesturing language that did not require changes to the larynx? Arbib has presented a theory of language evolution but has omitted to discuss the selection pressures involved.

According to the Machiavellian intelligence hypothesis (Byrne & Whiten 1988; Whiten & Byrne 1997), the human brain evolved because of the selection pressure to develop cognitive capacities that facilitate social manoeuvring. This would also suggest that language evolved through the need to communicate mental states. The evolution of language would be driven primarily by the need to discuss matters such as loyalty, betrayal, forgiveness, and revenge. Arbib uses few examples to illustrate the content of the language he is discussing; he mentions gestures used to describe flying birds, hunters directing each other, the tastes of food, and the use of fire to cook meat. His argument seems to assume that speech and gesture are used to discuss the physical activities of daily living, rather than to express feelings, desires, or intentions, or to consider the thoughts of conspecifics.

Also, Arbib derives his model of imitation from that proposed by Byrne and Russon (1998) following their observations of leaf-folding by mountain gorillas. This is an imitative task that requires replicating the structural organisation of an action, rather than the mental states driving it. Communicating the knowledge inherent to this skill is a relatively straightforward matter using action demonstration, whereas to describe it using only speech would be more difficult. Conversely, communication concerning invisible mental states may lend itself more to speech than descriptive gesture. Consider for example, "John wrongly thinks that Bob is jealous of me," or, "you distract John whilst I plot revenge against Bob." It may be that in the discussion of invisible mental states, speech can add a valuable modality of communication, which may even supplant manual and facial gesture.

Arbib does not mention the possible role of the mirror-neuron system in mentalising, or the importance of this mentalising function in imitation. Imitation involves incorporating a novel action into a pre-existing behavioural repertoire (Whiten et al. 2004). It follows that for this to occur, the observed behaviour must be compared with the existing knowledge of the behaviour. Therefore, imitation requires more than remembering and then replicating the components and organisational structure of an action sequence. Rather, imitation requires that the observer draw on his or her own knowledge of an action exhibited by a model. This includes the observer's knowledge of the action's relationships to causes, beliefs, goals, desires, effects, and agency. Only then can the observer understand the role of the action in the model's behaviour.

Actions are therefore vehicles for the thoughts that shape them, in that thoughts are carried by actions from mind to mind. Both imitation and "simulation theory of mind" involve observing actions or behaviours from a stance of using self-knowledge to predict the mental states behind them (Meltzoff & Decety 2003). This means that both "theory of mind" and imitation depend on relating perceived actions to their motor counterparts (Meltzoff & Prinz 2002). The mirror-neuron system is the prime candidate to serve this function (Gallese & Goldman 1998), not as the only component, but by providing the original action-perception links that constitute the evolutionary origins and the developmental core for social cognitive growth. I suggest that it is the capacity of the mirror-neuron system to represent an observed action as if it were the behaviour associated with a self-generated mental state, thereby allowing for attribution of intention (and a secondary representational capacity; see Suddendorf & Whiten [2001]), rather than its capacity for coding an action's organisational structure, which enabled the mirror-neuron system to serve highly flexible imitation and praxis.

The neurodevelopmental disorder of autism is characterised by major developmental impairment of social cognitive ability, including imitative and mentalising abilities. Another characteristic feature, that is highly discriminative diagnostically, is the reduced use of all gestures, whether descriptive, instrumental, emphatic, or facial (Lord et al. 2000). This suggests that the neural system in humans serving gestural communication is knitted to that serving other social cognition (Williams et al. 2001). Whether dysfunctional mirror-neuron systems account for this symptom cluster is still a matter for research, but it seems unlikely that during evolution, language became more divorced from social cognitive systems once it became spoken. Indeed spoken language can become divorced from social cognition in autism, when it may be repetitive, stereotyped, and pragmatically impaired, such that its communicative function is severely impaired. If language did evolve only as Arbib describes, it could be impaired in a similar manner.

I suggest that the evolution of language from object-directed imitation would have been intimately tied to the evolution of social communication at the neural level. During early hominid evolution, the representations being pantomimed through gestural communication (including facial expression) would have been concerned with mental states, including feelings and desires. Facial and manual gestures were being used by individuals to express both their own feelings and what they thought others were feeling. The neural systems serving these functions would form the basis for the communication of more complex mental states, which would recruit vocal and auditory systems as well as semantic and planning structures in temporal and frontal lobes.

In summary, I suggest that mirror neurons first evolved within social cognitive neural systems to serve a mentalising function that was crucial to their praxic role in imitation and gestural communication. As the evolution of social language was driven through the need to convey and discuss invisible mental states, and these became increasingly complex, so a vocal-auditory modality became recruited as an increasingly valuable additional means of communication. This extended, rather than altered, the fundamentally social nature and function of language, and maintained its dependence upon social cognitive mechanisms such as secondary representation.

ACKNOWLEDGMENTS
I am grateful to Thomas Suddendorf and Nina Dick-Williams for comments on an earlier draft of this commentary.

The explanatory advantages of the holistic protolanguage model: The case of linguistic irregularity

Alison Wray
Centre for Language and Communication Research, Cardiff University, Cardiff, CF10 3XB, Wales, United Kingdom. **wraya@cf.ac.uk**
http://www.cf.ac.uk/encap/staff/wray.html

Abstract: Our tolerance for, and promotion of, linguistic irregularity is a key arbitrator between Arbib's proposal that holistic protolanguage preceded culturally imposed compositionality, and the standard view that discrete units with word-like properties came first. The former, coupled with needs-only analysis, neatly accounts for the second-order linguistic complexity that is rationalised as fuzzy grammaticality, subclass exception, and full irregularity.

Any model of language evolution must explain four basic things:

1. The interface between real-world semantics and the arbitrary phonetic medium: a difficult problem, particularly if subcortical reflex vocalisations are not the precursor of speech;
2. The capacity for fast and fluent formulations of phonological strings, since this has no obvious purpose beyond language itself (unless for display);
3. Our ability to express and understand messages that juxtapose many separate meaning features; and
4. Why languages appear to be unnecessarily complex, relative to the perceived underlying simple rule systems.

Arbib's integrated model offers an explanation for the first three by identifying manual dexterity and imitation, exapted for pantomimic communication, as the conduit between holistic message and oral articulation. Associating Broca's area first with grasping and imitation is much more satisfactory than attributing to it an a priori involvement in language that must then be independently explained. Indeed, in line with Arbib's section 8, neurolinguistic and clinical evidence strongly suggests that linguistic representation in the brain is mapped on the principle of functional motivation, so language operations are expected to be distributed according to their primary functions or derivation (Wray 2002a, Ch. 14).[1]

However, Arbib's model also indirectly offers an explanation for point 4. In Arbib's scenario, complex meaning existed in holistic expressions before there was a way of isolating and recombining units. The subsequent application of what Arbib terms "fractionation" ("segmentation" for Peters [1983], who identified the process in first language acquisition) is viewed as culturally rather than biologically determined, and consequently, piecemeal and circumstantial rather than uniform and universal.

On what basis should we favour this proposal over the standard alternative (e.g., Bickerton 1996), that there have always been discrete units with word-like properties, which became combinable to create meaning, first agrammatically (protolanguage) and later grammatically? First, we can note that attributing to our biologically modern ancestors a default capacity for holistic rather than compositional expression, begs the question: Where is that holistic foundation now? Wray (2002a) demonstrates that holistic processing, far from being peripheral and inconsequential, is in fact alive and well and motivating much of our everyday linguistic behaviour.[2]

But I want to focus mainly on one linguistic phenomenon that has long caused puzzlement and demanded much explanatory effort: irregularity. It is surely a necessary corollary of the standard view of language as an ab initio combinatory system that we are predisposed to orderliness, and that unnecessary complexity and irregularity are an aberrance to be minimised rather than promoted or protected. Hence, first, we should find that languages attempt to cleanse themselves of phonological and morphological exceptions, oddities in patterns of lexical collocation, grammatical restrictions of the sort that demand subcategorisations of word classes, and lexical gaps. For instance, we would expect the upgrading of adjective subsets that cannot occur predicatively (**The objection is principal*) and attributively (**the asleep boy*), and the filling of gaps in lexical sets, for example, *horror/horrid/horrify, terror/*terrid/terrify, candor/candid/*candify* (Chomsky 1965, p. 186). Such cleansing does not generally occur. Most irregularity is preserved intact from one generation to the next. Although regularisation does happen at the margins, it is balanced by the creation of new irregularities (see below).

Second, children acquiring an L1 that is fully regular and transparent, such as Esperanto, ought to do so efficiently and perfectly. However, they do not (Bergen 2001). Instead, they introduce (apparently permanently) irregularities and sub-patterns that render complex the simple system of the input.

Third, if native speakers naturally develop a full compositional linguistic system during first language acquisition, we should expect their writing to reflect that system from the start. This, too, is not the case. In semi-literates, Fairman (e.g., 2000) reports *taket* (take it), *in form* (inform), *a quaint* (acquaint) and *B four* (before). Guillaume (1927/1973) offers *semy* (c'est mis), *a bitant* (habitant), *a ses* (assez) and *dé colle* (d'école). Thus is speech transcribed with strikingly little awareness of the grammatical or morphological components that are supposedly being freely manipulated.

All of these oddities are readily explained if humans are predisposed to treat input and output holistically where they can, and to engage in linguistic analysis only to the extent demanded by expressive need (rather than a principle of system) – *needs-only analysis* (NOA; Wray 2002a, pp. 130–32). Coupled with a parsimonious approach to pattern identification, NOA will:

a) Prevent the deconstruction of linguistic material that is no longer morphologically active, thus preserving irregularity;

b) Fence off combinations that are regular but are not observed to be subject to paradigmatic variation, and maintain them as complete units that cannot be generalised to other cases (as with the L1 acquisition of Esperanto); in so doing, protect the units from subsequent linguistic change, so they drift over time through fuzzy semi-regularity to full irregularity;

c) Support, in those who do not subsequently augment their fuzzy, half-formed linguistic system with formal training through literacy, a tolerance for underspecification and an absence of any expectation that language is *fully* composed of atomic lexical units. The bizarre spellings of semi-literates reflect a direct link between the whole meaning and its phonological form.

In addition, the fractionation of a holistic expression may often result in a "remainder" of phonological material that cannot be attributed a plausible meaning or function. Yet, because of (a) and (c), there may well never be a point when that material demands rationalisation – until the grammarian attempts to explain it in terms of a system it actually stands outside. Unless by haphazard or imposed hypercorrection, such irregular remainders may never be expunged and, although vulnerable to certain kinds of change, may persist in the long term, to the puzzlement of analysts (Wray 2002a) and frustration of adult language learners (Wray 2004).

Therefore, I contend that linguistic irregularity is a source of support for Arbib's proposal that compositionality is a choice rather than a fundamental in human language, and that its application is variable not absolute. Some aspects of what syntacticians are obliged to account for via complex rules may be no more than detritus from the process of fractionising unprincipled phonological strings.

If this is so, our challenge, before all the endangered languages disappear, is to recast our assumptions about prehistorical norms, by establishing what the "natural" balance is between compositionality and formulaicity in the absence of literacy and formal education. Many "fundamentals," such as the word, full classificatory potential, and inherent regularly of pattern, may come down to culture-centricity (Grace 2002) and the long-standing uneasy attempt to squeeze square pegs into the round holes of prevailing linguistic theory.

NOTES

1. This position easily supports Arbib's hypothesis (sect. 1.2) that there would be an extralinguistic human correlate of the primate mirror system for subcortical reflex vocalisations.

2. It was on the basis of this evidence that I first proposed a holistic protolanguage (Wray 1998; 2000; 2002b), but we avoid circularity since Arbib does not in any sense build his own story *upon* my proposal, he only cites it as an independently developed account consistent with his own.

Language evolution: Body of evidence?

Chen Yu[a] and Dana H. Ballard[b]

[a]*Department of Psychology and Cognitive Science Program, Indiana University, Bloomington, IN 47405;* [b]*Department of Computer Science, University of Rochester, Rochester, NY 14627.* **chenyu@indiana.edu dana@cs.rochester.edu http://www.indiana.edu/~dll/ http://www.cs.rochester.edu/~dana/**

Abstract: Our computational studies of infant language learning estimate the inherent difficulty of Arbib's proposal. We show that body language provides a strikingly helpful scaffold for learning language that may be necessary but not sufficient, given the absence of sophisticated language in other species. The extraordinary language abilities of *Homo sapiens* must have evolved from other pressures, such as sexual selection.

Arbib's article provides a complete framework showing how humans, but not monkeys, have language-ready brains. A centerpiece in hominid language evolution is based on the recognition and production of body movements, particularly hand movements, and their explicit representation in the brain, termed the mirror property.

How can we evaluate this proposal? One way is to take a look at infant language learning. The human infant has evolved to be language-ready, but nonetheless, examining the steps to competency in detail can shed light on the constraints that evolution had to deal with. In a manner similar to language evolution, the speaker (language teacher) and the listener (language learner) need to share the meanings of words in a language during language acquisition. A central issue in human word learning is the mapping problem – how to discover correct word-meaning pairs from multiple co-occurrences between words and things in an environment, which is termed reference uncertainty by Quine (1960). Our work in Yu et al. (2003) and Yu and Ballard (2004) shows that body movements play a crucial role in addressing the word-to-world mapping problem, and the body's momentary disposition in space can be used to infer referential intentions in speech.

By testing human subjects and comparing their performances in different learning conditions, we find that inference of speakers' intentions from their body movements, which we term embodied intentions, facilitates both word discovery and word-meaning association. In light of these empirical findings, we have developed a computational model that can identify the sound patterns of individual words from continuous speech using nonlinguistic contextual information and can employ body movements as deictic references to discover word-meaning associations. As a complementary study in language learning, we argue that one pivotal function of a language-ready brain is to utilize temporal correlations among language, perception, and action to bootstrap early word learning. Although language evolution and language acquisition are usually treated as different topics, the consistency of the findings from both Arbib's work and our work does show a strong link between body and language. Moreover, it suggests that the discoveries in language evolution and those in language acquisition can potentially provide some insightful thoughts to each other.

Language (even protolanguage) is about symbols, and those symbols must be grounded so that they can be used to refer to a class of objects, actions, or events. To tackle the evolutionary problem of the origins of language, Arbib argues that language readiness evolved as a multimodal system and supported intended communication. Our work confirms Arbib's hypothesis and shows that a language-ready brain is able to learn words by utilizing temporal synchrony between speech and referential body movements to infer referents in speech, which leads us to ask an intriguing question: How can the mirror system proposed by Arbib provide a neurological basis for a language learner to use body cues in language learning?

Our studies show quantitatively how body cues that signal intention could aid infant language learning. Such intentional body movements with accompanying visual information provide a natural learning environment for infants to facilitate linguistic processing. Audio, visual, and body movement data were collected simultaneously. The non-speech inputs of the learning system consisted of visual data, and head and hand positions in concert with eye gaze data. The possible meanings of spoken words were encoded in this nonlinguistic context, and the goal was to extract those meanings from raw sensory inputs. Our method first utilized eye and head movements as cues to estimate the speaker's focus of attention. At every attentional point in time, eye gaze was used as deictic reference (Ballard et al. 1997) to find the attentional object from all the objects in a scene, and each object was represented by a perceptual feature consisting of color, texture, and shape features. As a result, we obtained a temporal sequence of possible referents.

Next, a partitioning mechanism categorized spoken utterances represented by phoneme sequences into several meaning bins, and an expectation-maximization algorithm was employed to find the reliable associations of spoken words and their perceptually grounded meanings. Detailed descriptions of machine learning techniques can be obtained from Yu and Ballard (2004). The learning result is that this system can learn more than 85 percent of the correct word-meaning associations accurately, given that the word has been segmented. Considering that the system processes raw sensory data, and our learning method works in unsupervised mode without manually encoding any linguistic information, this level of performance is impressive.

Such results are very consistent with Arbib's proposal that these body constraints served to start language development on an evolutionary scale. However, this leaves unanswered the question of why *Homo sapiens evolved without language*. Arbib's argument seems to be that if a plausible sequence of steps is laid out, and the "height" or difficulty in transiting each step is small, then somehow evolution should have been compelled to follow this path. But our sequence of steps in the model of infant language learning also has small steps – recognize body movements, recognize intentions as communicated with body movements, recognize attentional objects in a scene, recognize the sounds that accompany these movements. These steps would be accessible for a variety of social species, and yet they were traversed only by humans.

Arbib makes special use of the hand representations, suggesting that perhaps humans had an edge in this category that provided the needed leverage. This is again very plausible, yet our studies show that you can get quite far just by hanging sounds on the end of the eye fixations and hand movements. From our point of view, any animal species that could communicate intention through body movement had the possibility of developing some kind of language. Hence, it is likely that some other constraints must be brought into play to account for the uniqueness of language in humans. Surprisingly, Arbib does not mention Miller's hypothesis that language is a product of sexual selection. Miller (2001) argues that the human brain must have been the kind of runaway process driven by sexual selection in a similar manner to Bower bird's nests and peacock's tails. Miller's arguments are extensively developed and show how *Homo sapiens* could have gotten a jump start on very similar species with very similar brain architectures.

Author's Response

The mirror system hypothesis stands but the framework is much enriched

Michael A. Arbib
Computer Science Department, Neuroscience Program and USC Brain Project, University of Southern California, Los Angeles, CA 90089-2520.
arbib@pollux.usc.edu http://www-hbp.usc.edu/

Abstract: Challenges for extending the mirror system hypothesis include mechanisms supporting planning, conversation, motivation, theory of mind, and prosody. Modeling remains relevant. Cospeech gestures show how manual gesture and speech intertwine, but more attention is needed to the auditory system and phonology. The holophrastic view of protolanguage is debated, along with semantics and the cultural basis of grammars. Anatomically separated regions may share an evolutionary history.

R1. Introduction

R1.1. The commentaries in perspective

The original mirror system hypothesis (MSH) states that:

> H1. The *parity requirement* for language in humans is met because Broca's area evolved atop the mirror system for grasping with its capacity to generate and recognize a set of actions.

The target article (TA) goes beyond MSH to distinguish a language-ready brain (equipping the child to learn a language) from a brain that "has" language (in the sense of, e.g., an innate "principles and parameters" universal grammar) and then to assert that:

> H2. Language readiness evolved as a multimodal manual/facial/vocal system with protosign providing the scaffolding for protospeech – these then co-evolved in an expanding spiral to provide "neural critical mass" for protolanguage

and further that:

> H3. Protolanguage was holophrastic – "protowords" were semantically more akin to phrases or sentences of modern language than words "as we know them."
>
> H4. Biological evolution gave humans a language-ready brain, but the emergence of human languages from protolanguage was a matter of history, not biology.
>
> H5. Whereas the original MSH focused on macaque F5 and Broca's area, F5 is part of a larger F5-PF-STS system in the macaque, and this "lifts" to a larger frontal-parietal-temporal language-ready system in the human brain.

Among them, H2 to H5 constitute an extended MSH. What needs stressing is that these four hypotheses are almost independent – and thus each must stand on its own. My response to the commentaries is grouped as follows:

Section R2 shows that complex imitation must be complemented by planning (R2.1) and viewed in developmental perspective (R2.2).

Section R3 generally endorses the role of the mirror system in evolution of the language-ready brain, but mechanisms supporting conversation (R3.1), motivation (R3.2), and theory of mind (R3.3) must also be taken into account.

Section R4 considers lessons from modeling biological neural networks (4.1) and evolving artificial networks (R4.2).

Section R5 reviews the debate over the claim (H2) that the path to protospeech was indirect. Discussion of co-speech gestures (R5.1) shows how strongly manual gesture and speech are intertwined. Future work must factor in new data on the auditory system (R5.2). Data on primate vocalization challenge H2 but do not topple it (R5.3). However, any claim that protosign had a large head start (if any) on protospeech in the expanding spiral is questionable. The challenge of evolving a phonological system remains (R5.4).

Section R6 discusses the transition from protolanguage to language with special attention to the debate on H3, the holophrastic view of language (R6.1). Issues on bringing semantics into MSH (R6.2) are followed by a brief discussion of H4, emphasizing the cultural basis of grammars (R6.3).

Section R7 revisits the overview in Figure 6 of the TA. Unfortunately, the figure was mentioned only in one commentary and then only in passing, but I discuss commentary relevant to the issues of whether anatomically separated regions may share an evolutionary history (R7.1) and how action planning supplements mirror systems in language evolution (R7.2).

In addition to the commentaries published here in this issue, I had the privilege of receiving equally fine commentaries from Yoonsuck Choe; Jean-Louis Dessalles & Laleh Ghadakpour; Peter Dominey; James Hurford; Masao Ito; David Kemmerer; Takaki Makino, Kotaro Hirayama & Kazuyuki Aihara (Makino et al.); Emese Nagy; Massimo Piattelli-Palmarini & Thomas Bever; Friedemann Pulvermüller; Andreas Rogalewski, Andreas Jansen, Ann-Freya Foerster, Stefan Knecht, & Caterina Breitenstein (Rogalewski et al.); Martin Ruchsow; Markus Werning; and Patricia Zukow-Goldring. These commentaries are posted on the BBSOnline Web site and have been given a fuller Author's Response. The supplemental commentaries with Author's Response are retrievable at the following URL: http://www.bbsonline.org/Preprints/Arbib-05012002/Supplemental/. I am particularly grateful to the many commentators whose correspondence allowed me to more fully understand the issues they raised. I cannot do justice to this "conversation" in 10,000 words here, but hope to develop many of the issues in *Beyond the Mirror: Biology and Culture in the Evolution of Brain and Language*, which I am currently preparing for publication by Oxford University Press.

I use **boldface** for commentators' names when responding to the present published commentaries and *italic* when discussing (more briefly) the supplemental ones. A commentator's name followed by the notation (p.c.) refers to the follow-up correspondence (personal communication), not the original commentary.

R1.2. Et cetera

A number of interesting points do not fit into the above framework:

R1.2.1. Birds and others. **Pepperberg** applies my criteria for language readiness to the behavior of the Grey parrot, Alex, that she has taught to communicate with humans using rudiments of English speech. Despite the lack of strong neural homologies between parrots, songbirds (Doupe & Kuhl 1999), and primates, we may still hope to model relevant circuitry (R4) to better understand what allows a neural network to achieve different language-related functions. **Fitch** notes that some species may have vocal but not bodily imitation, and vice versa. This is irrelevant to MSH, which asserts that humans had a particular history. This does not deny that comparative study of neural mechanisms underlying different forms of imitation may help us better understand the workings of the human brain – though the closer the homology, the more likely the payoff.

Pepperberg's assertion that little about my criteria for language readiness is unique to humans seems a blow to my claim to characterize what allows human children to learn "full language" where other species cannot. Perhaps there are differences of degree: for example, Alex does not meet my full criteria for "complex imitation." What enriches the discussion is that chimpanzees raised in a human environment can exhibit far more "protolanguage" than their wild cousins – observing animals in the wild does not define the limits of complexity of their behavior.

R1.2.2. Lateralization. **Kaplan & Iacoboni** show that motor activation to sight of an action is typically bilateral, whereas action sounds activate the motor cortex only in the left hemisphere. This may be related to evolutionary processes that lateralized language. Since lateralization has been debated extensively in *BBS* (Vol. 26, No. 2, Corballis 2003a), I will not comment here (but see R5.3) beyond the observation that, because children who receive a hemispherectomy early enough can gain fairly good command of language (though comprehension of syntax does show some left-hemisphere superiority [Dennis & Kohn 1975]), lateralization would seem to be not so much the genetic specification of different kinds of circuitry in the two hemispheres as a developmental bias which favors, but does not force, differential development of skills there.

R1.2.3. Sexual selection. **Yu & Ballard** cite the hypothesis that language is a product of sexual selection. I am unable to evaluate this hypothesis, but raise two questions: Does sexual selection function differently in early hominids and early great apes? Why does it not yield stronger dimorphism between male and female language use?

R1.2.4. Genetic underpinnings. **Théoret & Fecteau** note attempts to implicate the *FOXP2* gene in language. However, *FOXP2* is implicated in many systems, from the gut to the basal ganglia. It has been argued that because the gene changed only once from mouse to our common ancestor with the great apes but changed twice in the hominid line, it may hold the key to what distinguishes us from the great apes. However, the mutation of the gene seen in a number of members of the family KE does not reverse the two "recent" mutations to yield humans with a chimpanzee-like *FOXP2* gene. The KE language deficits seem more a function of motor problems than proving a causal relation between changes in *FOXP2* and the evolution of the language-ready brain (Corballis 2004). **Pepperberg**'s description of the use of expression of the *ZENK* gene to form a functional map of avian brains for behavior related both to auditory processing and vocal production and the coupling of this to neurophysiology, provides an encouraging model for future studies in macaques.

R2. Complex imitation

R2.1. Complex imitation and planning

I hypothesize that the mirror system for grasping evolved in two stages: first to provide feedback for dexterous manual control, then to underwrite the ability to act with other brain regions to make information available for interacting with others. *Makino, Hirayama & Aihara* (*Makino et al.*) observe that success in complex imitation requires the ability to recognize the goal of an action as the basis for mastering the action which achieves that goal. Indeed, Arbib and Rizzolatti (1997) gave the equation Action = Movement + Goal, and the mirror neurons system (MNS) model recognizes an action in terms of the goal of successful grasping of an affordance. Complex imitation takes us further. It rests on recognizing how different actions fit together to achieve various subgoals of the overall goal.

Bickerton and **Prudkov** assert that there cannot be imitation unless someone has first created something to imitate, and that mirror neurons offer no clue as to how totally novel sequences could have been created. Actually, new skills can emerge by trial and error. The problem is to preserve them. The data on chimpanzee cultures (Whiten et al. 2001) show how few skills chimpanzees acquire. I suggest that it is complex imitation that enables humans to move beyond such limited repertoires, cumulatively ratcheting up the available stock of novel skills.

Complex imitation presupposes a capacity for complex action analysis – the ability to analyze another's performance as a combination of actions (approximated by variants of) actions already in the repertoire. In modern humans, imitation undergirds the child's ability to acquire language, whereas complex action analysis is essential for the adult's ability to comprehend the novel assemblage of "articulatory gestures" that constitute each utterance of a language. However, the adult does not imitate this assemblage but rather factors it into the planning of his reply. I agree with **Bridgeman** that mirror systems must be supplemented by a planning capability to create, store, and execute plans for sequences of actions and communicatory acts. These apparent sequences are the expression of hierarchical structures. In Figure 5 of the TA, interpretation of actions of others is coupled to planning of one's own actions; Bridgeman stresses the need for the complementary evolution of these two capabilities. They underlie perception grammars and production grammars, mentioned in discussion of Figure 1 of the TA.

Bickerton observes that when someone addresses you, you do not just imitate what they said. True. The human mirror system creates a representation that can be used for feedback control, imitation (which monkeys do not exhibit), or generating some appropriate response while inhibiting mimicking. Only in pathology does this inhibition fail, yielding compulsive imitation (echopraxia; Podell et al. 2001).

R2.2. Imitation in developmental perspective

Zukow-Goldring sees affordances and effectivities (what the body can do; Shaw & Turvey 1981) as two sides of the mirror system. By directing the child's attention to its own effectivities in relation to affordances, the caregiver narrows the search space for learning, and thus enhances that learning (Zukow-Goldring 1996). These practices may pave the way to early word learning. The prolonged period of infant dependency in humans combines with caregiving to provide conditions for complex social learning.

Neonatal imitation is based on moving single effectors and thus differs from goal-directed imitation. (Studdert-Kennedy [2002] discusses data consistent with the view that the infant at first imitates sounds by moving one articulator and only later coordinates articulators.) Social reciprocity in neonatal imitation (R3.1) may be a necessary precursor for complex imitation, establishing that "I am like the other." Biological evolution may have selected for neonatal imitation as a basis for complex imitation.

Yu & Ballard found that body cues signaling intention can aid word learning in adults, suggesting the utility of such cues for children. Their computational model reliably associates spoken words and their perceptually grounded meanings. This model employs "small steps" which, they suggest, would be accessible by a variety of social species, and "yet they were only traversed by us." However, they seem accessible to parrots and bonobos as well as the 2-year-old child – which is why I emphasize complex imitation.

R3. Complementing complex imitation: Motivation and theory of mind

Conversation, motivation, and theory of mind – and prosody (R5.3) – must all be addressed in a satisfactory account of the language-ready brain. This requires expanding MSH rather than weakening it.

R3.1. Conversation

Kotchoubey emphasizes pragmatics, for example, what we say depends on the mental state (R3.3) of our "hearer." However, his claim that "We do not use language to transmit information, but to persuade and motivate" (R3.2) seems a false dichotomy. "Look at this beautiful flower" combines information – "This flower is beautiful" – and persuasion – "Look at this flower." Kotchoubey (personal communication) stresses that his starting point is cooperation between two or more humans, reinforcing the claims of MSH for relating praxic and communicative actions.

Nagy suggests an innate basis for conversation that precedes its pragmatic function – newborn infants communicate by using "imitation" right after birth (Nagy & Molnar 2004). She suggests that language develops from these early intersubjective "conversations" (Trevarthen 2001). The cycle of turn taking in "imitating" a small repertoire of "almost innate" gestures is crucial in establishing the social pattern of turn taking (R2.2). (Cf. "motherese"; R5.3.)

R3.2. Motivation

Prudkov downplays complex imitation, arguing that the complexity of languages builds on the ability of the human brain to construct diverse goals. He suggests that animals can form learned motivations only when basic drives are activated. However, animals can acquire secondary reinforcers, and so on. Chimpanzees have the ability to develop non-innate subgoals (e.g., cracking nuts). The mirror system is well linked to the motivational system in the macaque. The target article shows that the F5 mirror system for grasping is best understood within the larger F5-PF-STSa mirror system for manual and orofacial actions. Rizzolatti et al. (2001) observe that STSa is also part of a circuit that includes the amygdala and the orbitofrontal cortex and so may be involved in the elaboration of affective aspects of social behavior. Hence, Prudkov's transition to "non-innate motivation" may be less proximate for the evolution of the language-ready brain per se than complex imitation, which made possible the rapid acquisition of new skills.

R3.3. Theory of mind

Fabrega stresses that successful pantomime presupposes social cognition, awareness of self, and goal-setting – reversing the view of those who attribute self-consciousness to language (Macphail 2000). Fabrega (personal communication) also asks: "What are thoughts beyond internal use of language"? Arbib (2001a) suggests that there are forms of consciousness common to many mammals, but that mirror system developments help explain why humans also have forms of consciousness that build upon, rather than precede language. Development of increasing subtlety of language can feed back into the nonlanguage system to refine our perceptions and experience.

Fabrega says that I do not specify how much of the protosign/protospeech spiral is enough to support the cultural evolution of language, and he asks whether "protoculture" emerges as the expanding spiral gets underway. I suggest that chimpanzees have protoculture (Whiten et al. 2001), but that human culture is qualitatively different, and language makes it so.

Williams sees the greatest evolutionary advantage conferred by spoken language as its ability to communicate mentalistic concepts (theory of mind, ToM). Williams stresses selection pressure for social maneuvering where I have emphasized physical activities. Surely the two "domains of discourse" complement each other. Williams notes the possible role of the mirror neuron system in mentalizing (Gallese 2003; Meltzoff & Decety 2003). We need to investigate whether an account can be given of a shared evolution of "mirror systems" suiting both ToM and complex imitation. I hypothesize that the ancestral mirror system for manual praxis was distinct from the putative mirror system for facial expression of emotion. The former would support pantomime and thence on to multimodal symbols; and then the availability of symbols could enrich the latter to yield rudiments of ToM.

Indurkhya sees the key to language evolution in an ability to see and feels things from another perspective and stresses the role of metaphor. Projection restructures the target by creating a new ontology for it; generalization of responsiveness of a mirror neuron may provide a novel ontology for objects and actions that can newly yield this activity. The TA shows that pantomime must be supplemented by conventional gestures to yield protosign. Within language itself, metaphor broadens our language by extending (proto)words to new contexts. In some cases, context is enough to recapture the shade of meaning. In others, one must invent terms which can express shadings specific to the new domain.

Indurkhya suggests that the ability to project one's self into other animals or objects might mark a crucial transition in hominid evolution. I think this notion is important. Much work on empathy emphasizes the similarities between self and other – but one must be able to maintain different models of other agents, adaptively going beyond what is held in common to imagine essential differences.

Williams and **Théoret & Fecteau** see autism as providing a window on the role of the mirror system in ToM and language. (Théoret & Fecteau add analysis of blindness.) Deficits in autism are prominent in speech associated with social communication, but praxic aspects of language are fairly well preserved. Perhaps what is affected is not so much language per se as the integration of this with affect and ToM. Interestingly, autistics may exhibit stereotypic mimicking (which monkeys do not have). Hence, it must be reiterated that a fully functional human mirror system inhibits mere repetition (echopraxia and echolalia) and instead relates the perception of perceived actions to the planning of an appropriate course of action.

R4. Lessons from modeling

R4.1. Biological neural networks

Horwitz, Husain, & Guenther (**Horwitz et al.**) note the importance of babbling in the development of spoken and sign languages. The Infant Learning to Grasp Model (ILGM; Oztop et al. 2004), mentioned briefly in the TA, is a theory of how "manual babbling" leads to an effective set of grasps. Arbib and Rizzolatti (1997) discussed the relevance of inverse and forward models to MSH, building on insights of Jordan and Rumelhart (1992) into vocal babbling. *Ito* and *Makino et al*. also stressed the importance of internal models; see Carr et al. (2003), Makino and Aihara (2003), Miall (2003), Wolpert et al. (2003), and Ito (2005).

Because *Piatelli-Palmarini & Bever* note the problem of determining similarity criteria for learning models, it is worth noting that the MNS and ILGM models have "innate" hand-related biases which enable them to acquire a range of grasps without having them built in. MNS input is the hand state relating the hand to the goal affordance of the object. ILGM acquires grasps whose visual manifestation MNS is to learn. ILGM has as its basis that the child reaches for a salient object, executes the grasp reflex if palmar contact is made, and builds a repertoire of grasps based on those which prove to be stable – stability supplies the reinforcement signal.

Dominey models the transformation between semantic structures and grammatical structures. He exploits the developmental analog of fractionation of holophrases to yield "words" which fill slots thus formed in the holophrase. Dominey et al. (2003) suggest that the resultant categorical distinction between function and content elements evolved first for sensory-motor function and then was exploited for phrasal-conceptual function. Dominey sees his modeling as consistent with Ullman's (2004) declarative/procedural model, in which the mental lexicon depends on temporal-lobe substrates of declarative memory, while mental grammar depends on a "procedural" network of frontal, basal-ganglia, parietal, and cerebellar structures supporting learning and execution of motor and cognitive skills.

Horwitz et al. model how learning an auditory target for each native language sound may occur via a mirror neuron system. Guenther (p.c.) notes that in this modeling, the perceptual system organizes largely independently of the motor system, whereas motor development relies very heavily on the auditory perceptual system.

Horwitz et al. emphasize the importance of combining neural modeling with neurophysiological and brain imaging data. Horwitz and Tagamets (2003) and Arbib et al. (1994) developed techniques for using models of primate neurophysiological data to predict and analyze results of human brain imaging. Arbib et al. (2000) analyze imitation of motor skills, relating human brain imaging to data on the macaque mirror system.

R4.2. Evolving artificial networks

We turn to using "simulated evolution" to obtain versions of initially unstructured networks whose parameters fit them to better perform or learn a class of actions. (Cangelosi & Parisi [2002] and Briscoe [2002] include papers using artificial evolution to probe constraints on language, but almost all the papers are far removed from neurobiology.)

Parisi, Borghi, Di Ferdinando & Tsiotas (**Parisi et al.**) support MSH through computer simulations and behavioral experiments with humans which suggest that seeing objects or processing words referring to objects automatically activates canonical actions that we perform on them. Parisi et al. (p.c.) point out that because actions are the only inter-individually accessible aspect of behavior, interpreting meanings in terms of actions might explain how meanings can be shared.

Borenstein & Ruppin evolve networks in which evolution of imitation promotes emergence of neural mirroring. However, primate data suggest that neural mirroring preceded imitation in human evolution. Borenstein (p.c.) responds that the key point is that in their system only the evolution of imitation was solicited, yet a mirror system emerged – suggesting that the link between imitation and mirroring may be universal.

Pulvermüller opts for specifying the putative neural circuits of syntax and recursion first, and thinking about possible evolution later. In his model (Pulvermüller 2002), words are represented by distributed cell assemblies whose cortical topographies reflect aspects of word meaning; these assemblies are formed by correlation learning and anatomical constraints. Syntactic rules emerge from the interplay between sequence detectors and general principles of neuronal dynamics. My concern is that this approach is so focused on symbols that it is ill suited to grounding an evolutionary approach to neurolinguistics.

Fagg and Arbib (1992) modeled the surprising speed with which monkeys could learn to associate a visual pattern and a motor response. This led us to distinguish Stage I from Stage II learning. Stage I may take months of shaping for the monkey to learn a general task like "associate a new visual pattern on this screen with the correct pull on the lever in front of you and you will get a reward." In Stage II, the monkey knows the task, and then takes only seven or so trials to stabilize the correct response to a novel visual pattern (Mitz et al. 1991). My concern with models using small neural networks is that the search space is so restricted that Stage I is no longer necessary. As one approaches the more biological models of R4.1 one must constrain evolutionary models to provide insights that link to the anatomy and neurophysiology of real brains. A key challenge for MSH-related modeling is to understand how to "evolve" from a brain for which Stage I learning can never yield the ability to learn language, to a human brain in which perhaps 2.5 years of learning is required for Stage I to make possible the "word explosion" which distinguishes the human infant from the chimpanzee.

R5. The path to protospeech

Several commentaries concerned H2: protosign provides scaffolding for protospeech. Those who take a "speech only" approach ignore the fact that language is multimodal. However, future work on MSH needs greater attention to the auditory system.

R5.1. Co-speech gestures

McNeill, Bertenthal, Cole & Gallagher (**McNeill et al.**) show that speech and "gesticulations" form a single system – a "gesticulation" (Kendon 1988) is a motion that embodies a meaning relatable to the accompanying speech. About 90% of gesticulations synchronize with the speech segments with which they are co-expressive. I disagree with McNeill et al.'s claim that gesture and speech must convey the same idea unit. Kita and Özyürek (2003) compared speech-gesture coordination in Turkish, Japanese, and English descriptions of an animated cartoon. Gestures used to express motion events were influenced by how features of motion events were expressed in each language, but also by spatial information that was never verbalized. However, the key point is that gesticulation is truly part of language.

McNeill et al. reject the claim that language started as a gesture language that was supplanted by speech and stress the importance of a close coupling between manual and vocal action. However, they suggest that my concept of an expanding spiral of protosign and protospeech does not go far enough. They advocate the evolution of a speech-gesture system in which speech and gesture evolved in lockstep. This criticism may be mistaken. Gesticulations are part of language, not protolanguage. By contrast, protosign may indeed have had a pantomimic base, with protosign scaffolding protospeech.

In any case, **McNeill et al.** establish that protolanguage was multimodal and that gesture was not "turned off" in evolution. Relating this to brain function, McNeill et al. offer the telling example of a man who can control his limb movements only through arduous visually guided attentional control, yet can still gesticulate while speaking even when he cannot see his hands. Kemmerer describes a brain-damaged subject, with intact semantic and grammatical knowledge of motion events, whose ability to retrieve the phonological forms of concrete nouns, action verbs, and spatial prepositions was severely impaired but whose ability to produce gestures with language-typical information packaging was mostly preserved (Kemmerer et al., submitted).

Emmorey concedes that the existence of modern sign languages might seem to support my hypothesis that there was an early stage in the evolution of language in which communication was predominantly gestural. However, she rejects this view because "the only modern communities in which a signed language is dominant have deaf members."

However, there are communities of hearing people using a signed language, albeit not their primary one (Kendon 1988). Emmorey suggests that sign languages can tentatively be traced back only 500 years, but such historical estimates are suspect. For example, http://www.ASLinfo.com/trivia.cfm says that by A.D. 530 Benedictine monks had invented signs to circumvent their vow of silence.

Emmorey asserts that "If communicative pantomime and protosign preceded protospeech, it is not clear why protosign simply did not evolve into sign language." **MacNeilage & Davis** suggest that I am "vulnerable" because I posit an open pantomimic protosign stage, whereas Hockett (1978) asserted that if manual communication had ever achieved openness, we would never have abandoned it for speech. However, I make no claim that protosign by itself realized the full potential of this openness. Emmorey further asserts: "A gestural-origins theory must explain why speech evolved at all, particularly when choking to death is a potential by-product of speech evolution." First, I do see slight advantages for speech over sign (agreeing with Corballis 2002) or, rather, for the early combination of protospeech with protosign over protosign alone, though this judgment is subjective. Second, Clegg and Aiello (2000) show that the risk of choking is highly overstated: "Mortality statistics for England & Wales . . . [show] that overall mortality from choking on food was very low averaging 0.6 per 100,000 head of population." Third, just as it is a matter of historical contingency that some tribes have both signed and spoken languages, it may well be that some tribes of early humans had languages dominated by speech and others had protolanguages dominated by sign. At a time of evolutionary bottleneck before humans left Africa 50,000 years ago, speech could have taken a dominant role. The counter-question to Emmorey is then: "If speech has primacy and sign is a modern innovation, how can one explain the ubiquity of co-speech gestures?"

R5.2. Taking the auditory system seriously

Bosman, López & Aboitiz (**Bosman et al.**), **Rauschecker**, and **Horwitz et al.** all make clear that future work on MSH must pay more attention to data on the auditory system than does the TA.

Bosman et al. (personal communication) argue that the biological principles that supported the evolution of mirror neurons for grasping may also have independently supported the evolution of auditory mirror neurons, but they agree that gesture may have helped establish certain semantic aspects of protolanguage by the use of pantomime. Their view is that protosign and protospeech coexisted and coevolved, and each contributed to the development of the other.

Bosman et al. discuss neurons in frontal areas of the monkey that respond strongly to vocalizations and thereby suggest that this domain may be the precursor of a vocalization mirror system similar to the mirror system for grasping. **Rauschecker** presents further relevant material on the macaque auditory system (R7).

Horwitz et al. note that a human may recognize 10^5 auditory objects, whereas the number of those that interest a monkey seems small. Moreover, monkeys seem far better in vision than audition in the use of long-term memory for objects. They thus argue that biological evolution gave hominids the ability to better discriminate and categorize auditory objects, retain them in memory, and relate them to articulatory gestures. I would agree, while noting that the success of speech has been linked to the ability to form an immense vocabulary from a small set of "phonemes" (the Particulate Principle; Studdert-Kennedy 2002).

R5.3. Primate vocalization

The issue is whether primate calls evolved directly to speech. **Seyfarth** argues that the parallels between primate same-different judgments for calls and children's perception of words cannot be ignored. However, such parallels suggest properties of auditory discrimination necessary for protospeech but do not help explain the crucial transition to production of an open-ended repertoire of symbols linked to an open semantics. Seyfarth faults me for characterizing primate vocalizations as "involuntary" signals but Note 6 of the TA addresses this explicitly. Seyfarth shows that I am wrong to deny that primate vocalizations are related to cortical regions, but his data primarily concern audition. **Bosman et al.** suggest that in the monkey there is overlap between area F5 and the cortical larynx representation, but **Gilissen** argues that monkey calls cannot be used as models for speech production because they are genetically determined in their acoustic structure. 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).

I have been unable to consult Cheney and Seyfarth's (in press) paper which apparently asserts that primate social knowledge bears striking resemblances to the meanings we express in language, which are built up by combining discrete-valued entities in a structured, hierarchical, rule-governed, and open-ended manner. Though uninformed, I speculate that **Seyfarth** may be misled by our human ability to offer a language-like description of the primates' abilities. This is not to deny that prelinguistic knowledge of social relations is relevant to evolving the language-ready brain (R3.3).

Provine discusses contagious yawning and laughing. These seem analogous to the contagious alarm calls of nonhuman primates. He observes that laughter is a ritualization of the sound of labored breathing in rough-and-tumble play – but, presumably, we are talking of biological selection rather than the free symbol creation to which ritualization contributes at the phonological and morphosyntactic level in language (Bybee 2001). Laughter punctuates the speech stream, in contrast with the tight integration of gesticulation and speech (**McNeill et al.**).

Kotchoubey and **Fitch** note that my emphasis on cognitive-symbolic aspects of language ignores prosody. Kotchoubey notes that prosody subserves both affective prosody (emotional expression) and linguistic prosody (as in distinguishing between an assertion and a question) and that both forms of prosodic information are processed mainly in the right temporal lobe. In similar vein, **Gilissen** notes that human vocal behavior does resemble monkey calls in the emotional intonations superimposed on the verbal component. Kotchoubey (p.c.) observes that in many languages, intonation is the only distinction between question and declaration. He thus suggests that linguistic prosody is a part of the right hemisphere so closely controlled by the left that they cannot work without each other. This is reminiscent of the coupling of gesticulations to the syntax and semantics of a specific language.

Gilissen cites Falk's (2004a) evolutionary perspective on the hypothesis that, as human infants develop, a special form of infant-directed speech (motherese) provides a scaffold for their eventual acquisition of language. This enriches our discussion of the role of the caregiver in neonatal "conversation" (R3.1). Gilissen says that the special vocalizations of human motherese are in marked contrast to the relatively silent mother/infant interactions that characterize chimpanzees, yet suggests a possible link between monkey calls and motherese. This apparent contradiction suggests that the affective content of motherese (and protolanguage) builds upon the monkey vocalization system, but the information content of motherese (and protolanguage) has a complementary evolutionary history. **Kotchoubey** suggests that the left-hemispheric subsystem develops as described by MSH to subserve the cognitive-symbolic function, whereas the right-hemispheric subsystem is a direct successor of monkey vocalization mechanisms and gives language its intonational color. It is a long-standing observation (Hughlings Jackson 1878–79) that imprecations survive damage to the human brain that blocks normal speech. In Arbib (2002), I therefore suggested that the language-ready brain integrates action-oriented and affect-oriented systems in a pattern of cooperative computation.

Fitch adopts Darwin's hypothesis that our prelinguistic ancestors possessed an intermediate "protolanguage" that was musical and that music scaffolds the early structural and imitative aspects of language (prosody). He sees the semantic stage as coming later. However, even if we accept the importance of "musicality," it does not follow that the coevolution of vocal and manual gesture is tied more closely to music than to pantomime and linguistic communication – but it does encourage us to investigate how dance and music might enrich MSH.

R5.4. Evolving a phonological system

MacNeilage & Davis argue that my view that pantomime could be an open system disregards the view that for language to become an open system it must have a combinatorial phonology consisting of meaningless elements. However, I explicitly distinguish pantomime from protosign. But I do say too little about getting from a pantomimic repertoire to a speech repertoire. The following, adapted from Arbib (in press), may be helpful:

Signing exploits the signer's rich praxic repertoire of arm and hand movements, and builds up vocabulary by language-sanctioned variations on this multi-dimensional theme (move a hand shape along a trajectory to a particular position while making appropriate facial gestures). By contrast, speech has no rich behavioral repertoire of non-speech movements to build upon. Instead evolution took a particulate path, so that the word is built (to a first approximation) from a language-specific stock of phonemes (actions defined by the coordinated movement of several articulators, but with only the goal of sounding right rather than conveying meaning in themselves). On this analysis, a basic reach and grasp corresponds to a single word in signed language; whereas in speech, a basic reach and grasp is akin to a phoneme, with a word being one level up the hierarchy. In either case, the brain must provide a computational medium in which already available elements can be composed to form new ones, irrespective of the level at which these elements were themselves defined.

I characterized MacNeilage's frame/content theory as being about the evolution of syllabification but offering no clue as to what might have linked such a process to the expression of meaning. **MacNeilage & Davis** note that they now address this criticism by arguing that the first words may have been kinship terms based on baby talk (MacNeilage & Davis, in press b – I received the final version only after the TA was "set in concrete"). I do not deny that words like "mama" and "dada" may have been based on baby talk. But to suggest that this gives us insights into the emergence of protolanguage seems to me to conflate phylogeny and ontogeny – the prototalk of adult hunter-gatherers is unlikely to have been much like baby talk.

For **Fabrega**, the complexities of speech production seem in excess of what protosign/protospeech spiraling entails. I disagree. Even a protovocabulary of a few hundred protowords would already provide selective advantage for changes in the vocal apparatus which "full" language could exploit without further change. In any case, I insist that the appropriate framework must also explain co-speech gestures.

Kaplan & Iacoboni argue that mirror neurons in premotor cortex that respond to the visual and auditory consequences of actions allow for a modality-independent and agent-independent coding of actions, which may have been important for the emergence of language. Kaplan and Iacoboni (in preparation) found that when subjects simultaneously saw and heard an action, there was greater activity in the premotor cortex compared with control conditions in which they only saw or only heard the action. *Rogalewski et al.* report the use of trans-cranial magnetic stimulation (TMS) to show that linguistic tasks, like speaking, covert reading, and listening to speech, activate the hand motor system bilaterally (Floel et al. 2003). Kaplan & Iacoboni argue that the role of audiovisual mirror neurons in the evolution of language deserves more attention. I agree, but suggest that the best framework for this is provided by the expanding spiral hypothesis. In discussing Kohler et al. (2002) and Ferrari et al. (2003), the TA argued that these data do not support the claim that protospeech mechanisms could have evolved from F5 without the scaffolding provided by protosign. This matter is further debated by Fogassi and Ferrari (in press), Arbib (2005), and MacNeilage & Davis (in press b).

R6. From protolanguage to language

R6.1. For and against holophrasis

The hypothesis that protolanguage was based on holophrases was offered as an alternative to the view of protolanguage utterances as strings of "words as we know them."

Fitch supports the holophrase theory of language origin but suggests that Baldwinian exaptations may underlie the first behavioral stages in the transition from holistic communication toward modern language. I accept that the development of an articulatory system adequate to the demands of (proto)language phonology may have involved a Baldwinian effect but doubt Fitch's claim that the transition to language must have been "strongly and consistently shaped by selection [. . .], given the communicative and conceptual advantages that a compositional, lexicalized language offers." Agriculture, writing, and living in cities provide evidence that being advantageous does not imply a genetic change. Because of this I took pains to make clear that one's account of the evolution of the human brain might be seen as having two very different results: "the language-ready brain" versus "the brain that 'has' language."

Bridgeman argues against holophrasis. He asserts that monkey calls can be paraphrased in one or two words, such as "leopard." However, the leopard call's meaning can be better approximated in English by the sentence: "There is a leopard nearby. Danger! Danger! Run up a tree to escape." To this he might respond, "It's only one word, because 'leopard' is enough to activate the whole thing." But once one moves to protolanguage, one may want to convey meanings like "There is a dead leopard. Let's feast upon it," and we clearly cannot use the alarm call as the word for leopard in this utterance. Bridgeman asserts that the generality of words is about the same in all languages and therefore constitutes a possibly biological "universal" of language. However, it is well known that a word in one language may require a phrase or more to translate into another language. I therefore maintain that the size of words is a result of a long history of building increasingly flexible languages.

Bickerton is so "language-centered" that he gives us little help in imaginatively recreating possible scenarios for a time when hominid protolanguage was at an early stage of development. He asserts that it is "questionable whether any species could isolate 'a situation' from the unbroken, ongoing stream of experience unless it already had a language with which to do so." But we know that biological evolution yielded a repertoire of primate calls each of which is akin to (but very different from) a "protoword" describing a "situation" in my sense, and I have tried to imagine how the brain could have so evolved that such protowords could be "invented" and disseminated in hominid communities. I suggest that early hominids very rarely created protowords for new situations. I require only a slow accretion of such nameable situations in each generation to build towards the critical mass that constituted protolanguage. Bickerton notes that those who play charades use "a large set of 'disambiguating signs' – stereotypic gestures for 'film title,' 'book title,' and so on." I concede that early hominids had no signs for these! But the crucial point is this: When I posit that there is a protoword for "The alpha male has killed a meat animal and now the tribe has a chance to feast together. Yum, yum!", I do not claim that (at first) there were protowords for all the variations, such as "The alpha male has killed a meat animal but it's too scrawny to eat. Woe is we." I think this point also addresses one half of **MacNeilage & Davis**'s dismissal of my supposed claim that hominids in the protospeech stage could have "dashed off complex semantic concepts with holistic phonetic utterances" (those are their words not mine). Bickerton cites Tallerman (2004), who argues that holophrasis was incompatible with contrastive phonology, but (as argued above) as the protovocabulary increased, the different protowords (whether signed, spoken, or both) would need to be readily generated and comprehended, and this could provide as much pressure for the particulate principle as does the anti-holophrase position.

Bickerton (p.c.) makes the telling point that I have offered no example of a hypothetical conversation consisting of representations of frequently occurring situations and that any model that will not do conversation (R3.1) is "worse than dubious." I have given too little thought to this, but suggest that protoconversations may have been like the interactions that we see in nonhuman primates, with a few protowords interspersed among actions, rather than taking – from the start – the form of a steady interchange of protowords.

I accept **Bickerton**'s argument that it is implausible that all "real words" are foreshadowed by widely distributed fragments of protowords. However, Kirby's (2000) computer simulation shows that statistical extraction of substrings whose meanings stabilize can yield surprisingly powerful results across many generations. I thus see the "Wray-Kirby mechanism" as part of the story of the protolanguage-language transition, but not the whole one. My sour fruit story, plus aspects of ritualization, provides other mechanisms whereby the virtues of a synthetic description might emerge – with the consequent demand for a protosyntax to disambiguate combinations once the combinatorics began to explode.

Wray notes that neurolinguistic and clinical evidence suggests that linguistic representation in the brain is mapped on the principle of functional motivation (Wray 2002a, Ch. 14). Wray (p.c.) expands on this as follows: When people lose language skills after a stroke, it is common for them to retain formulaic expressions such as "good morning" while they are unable to compose novel messages (cf. R5.3). She focuses on the functions of the material, and proposes that the functional locus supporting a class of lexical material – for example, names for people whose faces we recognize might be activated via the mechanisms that process visual face recognition, whereas expressions used for context-determined phatic interaction would be activated via, say, the right-hemisphere areas that handle context pragmatics – would be linked to the "language" areas of the left hemisphere. Damage to left-hemisphere language areas could block the ability to generate names and expressions on request, but spare the ability to use the words and expressions themselves, if activated as functional wholes.

Wray supports the holophrasis theory by focusing on the linguistic phenomenon of irregularity. She presents a number of "oddities" about language use that are readily explained if humans are predisposed to treat input and output holistically where they can, and to engage in linguistic analysis only to the extent demanded by expressive need. Her formulation bridges between "true wholes" (protolanguage holophrasis) and "apparent compounds" (formulas within a modern language) – supporting our view that the "protolexicon" had many such wholes, rather than combining "words as we know them." Wray shows that the "holophrastic impulse" remains part of modern language use, even though languages have long supplanted protolanguages in human society.

Bridgeman denies holophrasis, but then asks "How could the sorts of words that cannot be used alone get invented?" and looks for evidence to the development of language in children. He concedes that a child's first utterances are holophrases but "next comes a two-word slot grammar." But ontogeny does not equal phylogeny. The child "extracting statistics" from the adult word stream is a far cry from early hominids, for whom very few (proto)words already existed. In modeling the 2-year-old, Hill (1983; Arbib et al. 1987; cf. Peters 1983 for data) sees the child as extracting fragments of the adult's speech stream to provide a set of templates that describe situations. At first, "want milk" or "love Teddy" are unanalyzed wholes, but then words common to many templates crystallize out, and word categories follow as it is recognized that certain words can fill similar slots (recall *Dominey*'s modeling, R4.1). This supports the claim that "holophrasis" is prototypical but that modern communities provide a setting in which the "Wray-Kirby mechanism" can extract existing words in a few years rather than extending the protovocabulary over the course of many generations.

Piattelli-Palmarini & Bever note that although idioms like "he kicked the bucket" may be semantically non-compositional, they do obey strict syntactic constraints. However, it is a mistake to confuse Wray's observation of the role of formulas in modern language use with the idea that protowords were themselves formulas. We are trying to imagine hominids without syntax and understand how they did get syntax. *Hurford* asserts that a synthetic evolutionary route to compositional syntax is "simpler" than the analytic (Wray-Arbib) account. But by what measure? Once you have discovered the combinatorial power of using syntax to combine words, then words are simpler. But if you have not done so, then labeling significant events or constructs seems the simpler strategy – which is why I still support the holophrasis theory as a viable alternative to the synthetic account.

R6.2. Bringing in semantics

The issue of semantics was emphasized not only by **Indurkhya** (see sect. R3.1) but also by *Choe, Dessalles & Ghadakpour, Hurford, Pulvermüller, and Werning* (supplemental commentaries). Hurford caught me in an embarrassing lapse. When I disagreed with "Hurford's suggestion that there is a mirror system for all concepts," it turns out that I was disagreeing not with the ideas published in Hurford (2004) but with Hurford's preliminary ideas as expressed in early discussions before the paper was written. I am happy to report that we are now in essential agreement about the diversity of perceptual and motor schemas.

Choe cites a thought experiment of Choe and Bhamidipati (2004) to assert that voluntary action can provide meaning to one's internal perceptual state, and that maintained invariance in the internal perceptual state can serve as a criterion for learning the appropriate action sequence. **Parisi et al.** assert "If we interpret not only signs but also their meaning in terms of motor actions, we can understand how meanings can be shared between speakers and hearers. Motor actions are the only aspect of behavior which is inter-individually accessible." However, there is no "reddish action" as such. And we must be careful about jumping to the idea that every gesture mimics a direct action. Nonetheless, I welcome the discussion by Parisi et al. of evidence that language is grounded in action. Studies on the neural basis of cognition suggest that different areas are activated for manipulable and non-manipulable objects (Chao &Martin 2000); manual gestures may be automatically activated not only by visual stimuli but by words, too (Gentilucci 2003a; 2003b); and Borghi et al. (2004) found in a part verification task that responding by moving the arm in a direction incompatible with the part location was slow relative to responding in a direction compatible with the part location.

This could explain why *Ruchsow* argued that MSH is in good agreement with an externalistic account of semantics. The externalist denies that there is any fact about my mental or neural states that constitutes the meaning of my words (Kripke 1982; Wittgenstein 1958). For internalists, conversely, cognition is computation over (symbolic) representations (Kurthen 1992). Ruchsow rejects internalism because it lets us "throw the world away," allowing reason and thought to be focused on the inner model instead (Clark 1999). Ruchsow finds that many passages in the TA can be read in favor of externalism but sees "some sympathy for internalism" in references to fMRI and PET studies. Actually, I regard both externalism and internalism as incomplete and have sought a framework in which the partial truths of each can be integrated. Arbib and Hesse (1986) expanded my "internal" schema theory of "schemas in the head" to provide a complementary account of "external" schemas that reside in the statistics of social interaction and are thus the expression of socially shared (externalist) knowledge.

Werning confronts the "complex first" paradox: substance concepts are more frequently lexicalized across languages than attribute concepts, and their lexical expressions are ontogenetically acquired earlier. This is hard to reconcile with the view that prototypical substance concepts are semantically complex so that, for example, the substance concept [mango] is made up of the vector of attribute concepts <orange, oval, big, soft, sweet, edible, . . .>. My solution is to distinguish the distributed code for "orange" as a feature implicit in early visual processing from the neural representation of "orange" as a concept that can be put into words. Werning cites Fodor's (1995) view of mental representations to argue that it would be logically impossible to have two representations with the same content in one and the same brain. However, data reviewed in section 3.1 of the TA show that the size of an object may have different representations for grasping (dorsal) and "declaring" (ventral). Similarly, the color of an object may be salient in segmenting it from the background or distinguishing it from other objects, yet not enter consciousness. For the child, the redness of his truck is as indissoluble from the truck's identity as the fact that it has wheels – so the word "truck" may imply "redness" and "wheels" when the child does not have words or well-formed concepts for either.

Dessalles & Ghadakpour stress that an account of the evolutionary emergence of language should "explain why and how our ancestors got minds able to form predicative structures, and to express them through compositional languages." (See Hurford [2003] for a different approach to this problem.) They point to a crucial distinction between gaining the use of predicative structures to communicate some aspects of a situation and using predicates to "think new thoughts": "We can systematically express the negative version of a predicative structure, for example, 'Leo doesn't grasp the raisin,' whereas there is no perceptive meaning corresponding to the negation of a visual scene." I would agree, yet would suggest that one can imagine a series of "inventions" that would build more and more general power into such a capability. At first, the effort to apply certain recognition criteria failing to meet the threshold is what justifies the negation. This requires that context demands that only some tests, of all possible perceptual tests, be applied (cf. presuppositions). I would probably not say to you "There are no onions in my office" unless, for example, you knew I was in my office and had expressed your need of onions. Section 7 of the TA talks of language capabilities being extended by bricolage over a long period of time, and argues that "the language-ready brain" provided by the genome lacks much of what we now take for granted as parts of language. I view *and, not,* and *every* as human inventions augmenting language and thus reshaping thought.

R6.3. Concerning innate universal grammar

Kotchoubey questions my view that the development of language from protolanguage was social by noting that dissemination of "inventions" currently exploits such institutions as writing, hierarchical social organization, and mass media. I respond that it was because early humans did not have such institutions that their inventions would diffuse more slowly by protoword of mouth – and hand. That is why I think that the "cumulative bricolage" that led to the earliest "full" languages may have taken 100,000 years. Kotchoubey (p.c.) responds that the social mechanisms present from the very beginning, for example, socialization in tribes and education in families, are known to be very conservative and to brake progress rather to promote it. He thus argues that development of the first language rested on biological natural selection.

Kotchoubey notes that degrees of linguistic and genetic similarity between populations correlate, and that the transition from protolanguage to language may have covered 1,500 to 2,000 generations, and so he cannot understand why biological mechanisms should be denied during the evolution of the very first language. Yes, Cavalli-Sforza et al. (1996) attest that the length of isolation of a group yields correlated amounts of drift in both genetic structure and language structure, but there is no suggestion that the genetic changes are linked to the language changes. The counter to my hypothesis, then, would be to offer concrete proposals on how to shift the boundary from my set of criteria for protolanguage further into the language domain of syntax and compositional semantics. But how far? My firm conviction is that the genome does not specify a principles-and-parameters universal grammar, but I could accept that phonological expression of hierarchical structure might require a biological change not within my current criteria for language readiness. (Recall **Fitch** on Baldwinian exaptation, R6.1.)

Kemmerer supports the view that grammatical categories gradually emerged over hundreds of generations of historical language transmission and change. Linguists identify grammatical categories primarily by formal criteria, but the criteria used in some languages are either completely absent in others or are employed in ways that seem bizarre compared to English. For example, verbs are often marked for tense, aspect, mood, and transitivity, but some languages, such as Vietnamese, lack all such inflection; Makah, on the other hand, applies aspect and mood markers not only to words that are translated into English as verbs, but also to words that are translated into English as nouns or adjectives. Croft (e.g., 2001) addresses such quandaries by "construction grammar," seeking to identify the grammatical categories of individual languages according to the constructions unique to those languages. Of course, one may relate these to semantic and pragmatic prototypes: prototypical nouns specify objects and have referential functions, prototypical verbs specify actions and have predicative functions, and so on. Such considerations are very much consistent with the notion that languages (not Language-with-a-capital-L) evolved culturally through bricolage within many communities and diffusion across communities (Aikhenvald & Dixon 2002; Dixon 1997; Lass 1997; Ritt 2004).

R7. Towards a mirror-system based neurolinguistic model

Only one commentary touched on Figure 6 of the TA, but I gather here a number of comments relevant to the program it exemplified.

R7.1. Anatomically distinct regions may share an evolutionary history

Figure 6 of the TA offered a highly conceptual extension of the FARS model to include the mirror system for grasping and the language system evolved "atop" this. I see the various circuits as evolutionary cousins but do not require that the same circuitry subserves them. Given this, I am grateful for the review by **Barrett, Foundas & Heilman** (**Barrett et al.**) of functional and structural evidence supporting differential localization of the neuronal modules controlling limb praxis, speech and language, and emotional communication, but I am puzzled as to why they view these data as justifying rejection of an evolutionary relationship between the underlying mechanisms. Barrett et al. assert that the TA treats different forelimb gesture classes interchangeably, whereas, for example, it cites data from Corina et al. (1992a) which separate pantomime from signing.

There is no space here for a disquisition on mammalian brain evolution. Let me simply refer to Kaas (1993), Butler and Hodos (1996), Krubitzer (1998), and Striedter (2004) for support of the conclusion that increasing complexity of behavior is paralleled by increases in the overall size and number of functional subdivisions of neocortex and the complexity of internal organization of the subdivisions, and that reduplication of circuitry may form the basis for differential evolution of copies of a given system, with differing connectivities, and so on, to serve a variety of functions.

Barrett et al. usefully summarize data showing that in most humans, the left hemisphere may be dominant in the control of vocalization associated with propositional speech, but the right hemisphere often controls vocalization associated with emotional prosody, automatic speech, and singing. Moreover, **Kotchoubey** notes that although the right temporal lobe is critical for recognition of prosody (R5.3), prosodic aspects of language are also severely impaired in patients with lesions to orbitofrontal cortex (which has links to the mirror system, R3.2) and the corpus callosum. The latter, presumably, is related to integration between the two hemispheres.

Such data must be taken into account in building upon Figure 6 of the TA but do not contradict MSH at its current level of detail.

R7.2. Action planning complements mirror systems

I have already (in R2.1) applauded **Bridgeman**'s insistence that mirror systems must be supplemented by a planning capability to allow language to evolve. Interestingly, the design of Figure 6 of the TA was motivated in part by the work of Bridgeman. Recall the crucial role of inferotemporal cortex (IT) and prefrontal cortex (PFC) in modulating affordance selection in the FARS model. In the psychophysical experiments of Bridgeman et al. (1997; Bridgeman 1999), an observer sees a target in one of several possible positions, and a frame either centered before the observer or deviated left or right. Verbal judgments of the target position are altered by the background frame's position, but "jabbing" at the target never misses, regardless of the frame's position. The data demonstrate independent representations of visual space in the two systems, with the observer aware only of the spatial values in the cognitive (inferotemporal) system. The crucial point here is that communication must be based on the size estimate generated by IT, not that generated by posterior parietal cortex (PP). Thus, all three paths of Figure 6 of the TA are enriched by the prefrontal system, which combines current IT input with memory structures combining objects, actions, and relationships. (The figure just shows the path IT → DPLF as it may affect Wernicke's area.)

Bosman et al. disagree with the contrast (Arbib & Bota 2003) between the MSH theory presented in the TA as being "prospective" and the Aboitiz and García (1997) theory as being "retrospective." Our point was that Aboitiz and García focused on lexicon and syntax and looked at what might support them, without suggesting the intermediate stages that might have emerged through evolutionary pressures before language itself "appeared on the scene." As noted in the TA, these researchers emphasize working memory, whereas earlier work on MSH failed to do so. Hence the inclusion of working memories in Figure 6 of the TA. Further modeling must also take into account issues discussed in R4.1.

Figure 6 of the TA has auditory input only to area Tpt, whereas **Rauschecker** notes that auditory objects, including speech sounds, are identified in anterior superior temporal cortex (aST), which projects directly to inferior frontal regions and not along a posterior pathway, as classically assumed. He suggests that aST supports both the decoding of vocalizations in nonhuman primates and the decoding of human speech: "the conclusion is hard to escape that . . . nonhuman primate vocalizations are an evolutionary precursor to human speech sounds" (cf. discussion of **Seyfarth**, R5.3). However, brain imaging of users of sign language (Emmorey 2002) suggests that the brain regions constituting the perceptual and motor periphery differ between sign (parietal lobe in, manual-facial out) and speech (temporal lobe in; vocal-articulatory out), but that there are large overlap regions assumed to be responsible for syntactic and semantic processing at a level abstracted from the peripheral codes. This is the set of "new" regions which my theory begins to explain. Given this, I would not expect "protospeech" as it "spirally evolves" with "protosign" to invent a whole new periphery, but rather, to co-opt available resources. My impression would be that the auditory system of nonhuman primates is (almost) adequate for speech perception (note the restrictions reviewed by **Horwitz et al.**, R5.2) whereas (R5.4) the motor side needed immense changes to get separable control of vocal articulators to make the sounds of speech (as distinct from primate calls).

R8. Envoi

By way of conclusion, I simply invite the reader to return to section R1.1 of this response and assess my claim that, overall, the extended mirror system hypothesis is alive and well. I believe it has survived most of the criticism that would destroy its key claims, but that the commentaries provide many challenges for linking the evolving mirror system to other systems and for filling in many details that remain too sketchy or have not yet received due attention.

References

Letters "a" and "r" appearing before authors' initials refer to target article and response, respectively.

Aboitiz, F. & García V. R. (1997) The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. *Brain Research Reviews* 25:381–96. [arMAA, CB]

Aboitiz, F., García, R., Brunetti, E. & Bosman, C. (in press) The origin of Broca's area and its connections from an ancestral working memory network. In: *Broca's area,* ed. K. Amunts & Y. Grodzinsky. Oxford University Press. [CB]

Acardi, A. C. (2003) Is gestural communication more sophisticated than vocal communication in wild chimpanzees? *Behavioral and Brain Sciences* 26:210–11. [CB]

Adolphs, R., Damasio, H. & Tranel, D. (2002) Neural systems for recognition of emotional prosody: A 3-D lesion study. *Emotion* 2:23–51. [BK]

Aikhenvald, A. Y. & Dixon, R. M. W. (2002) *Areal diffusion and genetic inheritance.* Oxford University Press. [rMAA]

Albert, M. L., Goodglass, H., Helm, N. A., Rubens, A. B. & Alexander, M. P. (1981) *Clinical aspects of dysphagia.* Springer-Verlag. [AMB]

Alissandrakis, A., Nehaniv, C. L. & Dautenhahn, K. (2002) Imitation with ALICE: Learning to imitate corresponding actions across dissimilar embodiments. *IEEE Transactions on Systems, Man, and Cybernetics, Part A* 32(4):482–96. [BI]

American Psychiatric Association (1994) *Diagnostic and statistical manual of mental disorders,* 4th edition. American Psychiatric Association. [HT]

Arbib, M. A. (1981) Perceptual structures and distributed motor control. In: *Handbook of physiology, section 2: The nervous system, vol. II: Motor control, part 1,* ed. V. B. Brooks, pp. 1449–80. American Physiological Society. [aMAA]

(2001a) Co-evolution of human consciousness and language. In: *Cajal and consciousness: Scientific approaches to consciousness on the centennial of Ramón y Cajal's textura,* ed. Pedro C. Marijuan. *Annals of the New York Academy of Sciences* 929:195–220. [arMAA]

(2001b) Computational models of monkey mechanisms for the control of grasping: Grounding the mirror system hypothesis for the evolution of the language-ready brain. In: *Simulating the evolution of language,* ed. A. Cangelosi & D. Parisi. Springer-Verlag. [aMAA]

(2002) The mirror system, imitation, and the evolution of language. In: *Imitation in animals and artifacts,* ed. C. Nehaniv & K. Dautenhahn, pp. 229–80. MIT Press. [aMAA]

(2003) Schema theory. In: *The handbook of brain theory and neural networks,* 2nd edition, ed. M. A. Arbib, pp. 993–98. A Bradford Book/MIT Press. [aMAA]

(2004) How far is language beyond our grasp? A response to Hurford. In: *Evolution of communication systems: A comparative approach,* ed. D. K. Oller & U. Griebel, pp. 315–21, MIT Press. [aMAA]

(2005) Interweaving protosign and protospeech: Further developments beyond the mirror. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems* 6:145–71. [arMAA, PFM]

(in press) A sentence is to speech as what is to action? [A Contribution to the Special Issue on Integrative Models of Brocas Area and the Ventral Premotor Cortex]. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*. [rMAA]

Arbib, M. A. & Bota, M. (2003) Language evolution: Neural homologies and neuroinformatics. *Neural Networks* 16:1237–60. [arMAA, CB]

Arbib, M. A. & Hesse, M. B. (1986) *The construction of reality.* Cambridge University Press. [rMAA]

Arbib, M. A. & Rizzolatti, G. (1997) Neural expectations: A possible evolutionary path from manual skills to language. *Communication and Cognition* 29:393–423. [arMAA, PFM]

Arbib, M. A., Billard, A., Iacoboni, M. & Oztop, E. (2000) Synthetic brain imaging: Grasping, mirror neurons and imitation. *Neural Networks* 13:975–97. [rMAA]

Arbib, M. A., Bischoff, A., Fagg, A. H. & Grafton, S. T. (1994) Synthetic PET: Analyzing large-scale properties of neural networks. *Human Brain Mapping* 2:225–33. [rMAA]

Arbib, M. A., Caplan, D. & Marshall, J. C., eds. (1982) *Neural models of language processes.* Academic Press. [rMAA]

Arbib, M. A., Conklin, E. J. & Hill, J. C. (1987) *From schema theory to language.* Oxford University Press. [rMAA]

Arbib, M. A., Érdi, P. & Szentágothai, J. (1998) *Neural organization: Structure, function, and dynamics.* MIT Press. [rMAA]

Arbib, M. A., Fagg, A. H. & Grafton, S. T. (2003) Synthetic PET imaging for grasping: From primate neurophysiology to human behavior. In: *Exploratory analysis and data modeling in functional neuroimaging,* ed. F. T. Sommer & A. Wichert, pp. 232–50. MIT Press. [rMAA]

Arbib, M. A., Oztop, E. & Zukow-Goldring, P. (in press) Language and the mirror system: A perception/action based approach to communicative development. *Cognitie Creier Comportament/Cognitive Brain Behavior* [aMAA]

Armstrong, D. F., Stokoe, W. C. & Wilcox, S. E. (1995) *Gesture and the nature of language*. Cambridge University Press. [aMAA, DM]

Atran, S. & Norenzayan, A. (2004) Religion's evolutionary landscape: Counterintuition, commitment, compassion, and communion. *Behavioral and Brain Sciences* 27(6):713–30. [HF]

Avikainen, S., Wohlschläger, A., Liuhanen, S., Hanninen R. & Hari, R. (2003) Impaired mirror-image imitation in Asperger and high-functioning autistic subjects. *Current Biology* 13:339–41. [HT]

Aziz-Zadeh, L., Iacoboni, M., Zaidel, E., Wilson, S. & Mazziotta, J. (2004) Left hemisphere motor facilitation in response to manual action sounds. *European Journal of Neuroscience* 19(9):2609–12. [JTK]

Aziz-Zadeh, L., Maeda, F., Zaidel, E., Mazziotta, J. & Iacoboni, M. (2002) Lateralization in motor facilitation during action observation: A TMS study. *Experimental Brain Research* 144(1):127–31. [JTK]

Baddeley, A. D. (1986) *Working memory.* Clarendon Press. [HF]

Ballard, D. H., Hayhoe, M. M., Pook, P. K. & Rao, R. P. N. (1997) Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences* 20:1311–28. [CY]

Barkow, J. H. (1989) *Darwin, sex and status: Biological approaches to mind and culture.* University of Toronto Press. [HF]

Baron-Cohen, S. (2002) I am loved, therefore I think. Book review of *The cradle of thought* by R. P. Hobson. *Nature* 416:791–92. [HT]

Barrett, A. M., Crucian, G. P., Raymer, A. M. & Heilman, K. M. (1999) Spared comprehension of emotional prosody in a patient with global aphasia. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology* 12:117–20. [AMB]

Barrett, A. M., Dore, L. S., Hansell, K. A. & Heilman, K. M. (2002) Speaking while gesturing: The relationship between spontaneous speech and limb praxis. *Neurology* 58:499–500. [AMB]

Barrett, A. M., Schwartz, R. L., Raymer, A. L., Crucian, G. P., Rothi, L. J. G. & Heilman, K. M. (1998) Dyssynchronous apraxia: Failure to combine simultaneous preprogrammed movements. *Cognitive Neuropsychology* 15:685–703. [AMB]

Bass, A. H. & Baker, R. (1997) Phenotypic specification of hindbrain rhombomeres and the origins of rhythmic circuits in vertebrates. *Brain Behavior and Evolution* 50(Suppl.1):3–16. [EG]

Bass, A. H. & McKibben, J. R. (2003) Neural mechanisms and behaviors for acoustic communication in teleost fish. *Progress in Neurobiology* 69:1–26. [EG]

Bauer, G. & Johnson, C. M. (1994) Trained motor imitation by bottlenose dolphins (*Tursiops truncatus*). *Perceptual and Motor Skills* 79:1307–15. [BI]

Bergen, B. K. (2001) Nativization processes in L1 Esperanto. *Journal of Child Language* 28:575–95. [AW]

Bergman, T., Beehner, J., Cheney, D. L. & Seyfarth, R. M. (2003) Hierarchical classification by rank and kinship in baboons. *Science* 302:1234–36. [RMS]

Bickerton, D. (1990) *Language and species.* University of Chicago Press. [DB]

(1995) *Language and human behavior.* University of Washington Press. [aMAA, HF, WTF]

(1996) *Language and human behaviour, second edition.* University College London Press. [AW]

(2003) Symbol and structure: A comprehensive framework for language evolution. In: *Language evolution,* ed. M. H. Christiansen & S. Kirby, pp. 77–93. Oxford University Press. [DB, WTF]

Billard, A. (2000) Learning motor skills by imitation: A biologically inspired robotic model. *Cybernetics & Systems* 32:155–93. [EB]

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A. & Ward, B. D. (2004) Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience* 7:295–301. [JPR]

Binkofski, F., Buccino, G., Stephan, K. M., Rizzolatti, G., Seitz, R. J. & Freund, H.-J. (1999) A parieto-premotor network for object manipulation: Evidence from neuroimaging. *Experimental Brain Research* 128:210–13. [aMAA]

Bischoff-Grethe, A., Crowley, M. G. & Arbib, M. A. (2003) Movement inhibition and next sensory state prediction in basal ganglia. In: *The basal ganglia, vol. VI*, ed. A. M. Graybiel, M. R. Delong, & S. T. Kitai, pp. 267–77. Kluwer/Plenum. [aMAA]

Blumenthal, A., ed. and trans. (1970) *Language and psychology: Historical aspects of psycholinguistics.* John Wiley. [DM]

Boesch, C. & Boesch, H. (1983) Optimization of nut-cracking with natural hammers by wild chimpanzees. *Behavior* 83:265–86. [arMAA]

Bookheimer, S. (2002) Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience* 25:151–88. [JPR]

Borenstein, E. & Ruppin, E. (2004) Evolving imitating agents and the emergence of a neural mirror system. In: *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems,* ed. J. Pollack, M. Bedau, P. Husbands, T. Ikegami & R. A. Watson. MIT Press. [EB]

(2005) The evolution of imitation and mirror neurons in adaptive agents. *Cognitive Systems Research (Special issue on Epigenetic Robotics)* 6(3):229–42. [EB]

Borghi, A. M., Di Ferdinando, A. & Parisi, D. (2002) The role of perception and action in object categorization. In: *Connectionist models of cognition and perception,* ed. J. A. Bullinaria & W. Lowe, pp. 40–50. World Scientific. [DP]

(submitted) Objects, spatial compatibility and affordances: A connectionist study. [DP]

Borghi, A. M., Glenberg, A. M. & Kaschak, M. P. (2004) Putting words in perspective. *Memory and Cognition* 32:863–73. [rMAA, DP]

Borghi, A. M., Parisi, D. & Di Ferdinando, A. (2005) Action and hierarchical level of categories: A connectionist perspective. *Cognitive Systems Research* 6:99–110. [DP]

Bosman, C., García, R. & Aboitiz, F. (2004) FOXP2 and the language working memory system. *Trends in Cognitive Sciences* 8:251–52. [CB]

Bostanov, V. & Kotchoubey, B. (2004) Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology* 41:259–68. [BK]

Boyer, P. (1994) *The naturalness of religious ideas: A cognitive theory of religion.* University of California Press. [HF]

Brauth, S. E., Liang, W. & Roberts, T. F. (2001) Projections of the oval nucleus of the hyperstriatum ventrale in the budgerigar: Relationships with the auditory system. *Journal of Comparative Neurology* 432:481–511. [IMP]

Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D. & Blumberg, B. (2005) Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial life* 11(1–2):31–62. [BI]

Bridgeman, B. (1992) On the evolution of consciousness and language. *Psycoloquy* 3(15) Consciousness (1). Refereed electronic journal, available at: http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?3.15. [BB]

(1999) Separate representations of visual space for perception and visually guided behavior. In: *Cognitive contributions to the perception of spatial and temporal events,* ed. G. Aschersleben, T. Bachmann & J. Müsseler, pp. 3–13. Elsevier Science. [rMAA]

(2003) *Psychology and evolution: The origins of mind.* Sage. [BB]

Bridgeman, B., Peery, S. & Anand, S. (1997) Interaction of cognitive and sensorimotor maps of visual space. *Perception & Psychophysics* 59:456–69. [rMAA]

Briscoe, T., ed. (2002) *Linguistic evolution through language acquisition: Formal and computational models.* Cambridge University Press. [rMAA]

Brown, R., Hobson, R. P., Lee, A. & Stevenson, J. (1997) Are there "autistic-like" features in congenitally blind children? *Journal of Child Psychology and Psychiatry* 38:693–703. [HT]

Burgess, N., Jeffery, K. F. & O'Keefe, J., eds. (1999) *The hippocampal and parietal foundations of spatial cognition.* Oxford University Press. [aMAA]

Butler, A. B. & Hodos, W. (1996) *Comparative vertebrate neuroanatomy: Evolution and adaptation.* Wiley. [rMAA]

Bybee, J. (2001) *Phonology and language use.* Cambridge University Press. [rMAA]

Byrne, R. W. (2003) Imitation as behaviour parsing. *Philosophical Transactions of the Royal Society of London. Series B* 358:529–36. [aMAA]

Byrne, R. W. & Russon, A. E. (1998) Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences* 21(5):667–84. [JHGW]

Byrne, R. W. & Whiten A. (1988) *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans.* Oxford University Press. [JHGW]

Caldwell, C.A. & Whiten, A. (2002) Evolutionary perspectives on imitation: Is a comparative psychology of social learning possible? *Animal Cognition* 5(4):193–208. [BI]

Cangelosi, A. & Parisi, D., eds. (2002) *Simulating the evolution of language.* Springer. [rMAA, DP]

Carey, D. P., Perrett, D. I. & Oram, M. W. (1997) Recognizing, understanding, and producing action. In: *Handbook of neuropsychology: Action and cognition, vol. 11,* ed. M. Jeannerod & J. Grafman, pp. 111–130. Elsevier. [aMAA]

Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C. & Lenzi, G. L. (2003) Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences USA* 100(9):5497–502. [rMAA]

Carruthers, P. & Boucher, J. (1998) Introduction: Opening up options. In: *Language and thought: Interdisciplinary themes,* ed. P. Carruthers & J. Boucher, pp. 1–18. Cambridge University Press. [HF]

Casey, S. K. (2003) "Agreement" in gestures and signed languages: The use of directionality to indicate referents involved in actions. Ph.D. Dissertation in Linguistics, University of California at San Diego. [aMAA]

Cavada, C. & Goldman-Rakic, P. S. (1989) Posterior parietal cortex in rhesus monkey: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *Journal of Comparative Neurology* 287:422–45. [aMAA]

Cavalli-Sforza, L. (1996) *Geni, popoli e lingue.* Adelphi. (German edition: *Gene, Völker und Sprachen*, trans. G. Memmert. Carl Hanser Verlag). [BK]

Cavalli-Sforza, L. & Feldman, M. (1981) *Cultural transmission and evolution.* Princeton University Press. [BK]

Cavalli-Sforza, L., Menozzi, P. & Piazza, A. (1996) *The history and geography of human genes.* (Abridged paperback edition). Princeton University Press. [rMAA]

Chao, L. L. & Martin, A. (2000) Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* 12:478–84. [rMAA]

Cheney, D. L. & Seyfarth, R. M. (1982) How vervet monkeys perceive their grunts: Field playback experiments. *Animal Behaviour* 30:739–51. [RMS]

(1988) Assessment of meaning and the detection of unreliable signals in vervet monkeys. *Animal Behaviour* 36:477–86. [RMS]

(1990) *How monkeys see the world: Inside the mind of another species.* University of Chicago Press. [RMS]

(1998) Why monkeys don't have language. *The Tanner Lectures on Human Values* 19:175–219. [RMS]

(in press) Constraints and preadaptations in the earliest stages of language evolution. *Linguistic Review*. [rMAA, RMS]

Choe, Y. & Bhamidipati, S. K. (2004) Autonomous acquisition of the meaning of sensory states through sensory-invariance driven action. In: *Biologically inspired approaches to advanced information technology,* ed. A. J. Ijspeert & M. Murata. Springer. [rMAA]

Chomsky, C. (1969) *The acquisition of syntax in children from 5 to 10*. MIT Press. [aMAA]

Chomsky, N. (1965) *Aspects of the theory of syntax.* MIT Press. [AW]

Clark, A. (1999) An embodied cognitive science? *Trends in Cognitive Sciences* 3:345–51. [rMAA]

Clayton, N. S., Bussey, T. J. & Dickinson, A. (2003) Can animals recall the past and plan for the future? *Nature Reviews Neuroscience* 4:685–91. [IMP]

Clegg, M. & Aiello, L. C. (2000) Paying the price of speech? An analysis of mortality statistics for choking on food. *American Journal of Physical Anthropology* (Supplement) 30:126. [rMAA]

Cole, J., Gallagher, S. & McNeill, D. (2002) Gesture following deafferentation: A phenomenologically informed experimental study. *Phenomenology and the Cognitive Sciences* 1:49–67. [DM]

Corballis, M. C. (2002). *From hand to mouth: The origins of language.* Princeton University Press. [rMAA, KE, DM]

(2003a) From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences* 26(2):199–260. [rMAA, CB, PFM]

(2003b) Laterality and human speciation. In: *Speciation of modern* Homo sapiens, ed. T. J. Crow. *Proceedings of the British Academy* 106:137–52. [aMAA]

(2004) FOXP2 and the mirror system. *Trends in Cognitive Sciences* 8:95–96. [rMAA, HT]

Corina, D. P., Poizner, H., Bellugi, U., Feinberg, T., Dowd, D. & O'Grady-Batch, L. (1992a) Dissociation between linguistic and nonlinguistic gestural systems: A case for compositionality. *Brain and Language* 43(3):414–47. [arMAA]

Corina, D. P., Vaid, J. & Bellugi, U. (1992b) The linguistic basis of left hemisphere specialization. *Science* 255:1258–60. [aMAA]

Corina, D. P., Jose-Robertson, L. S., Guillemin, A. High, J & Braun, A. R. (2003) Language lateralization in a bimanual language. *Journal of Cognitive Neuroscience* 15(5):718–30. [aMAA]

Coulmas, F. (2003) *Writing systems: An introduction to their linguistic analysis*. Cambridge University Press. [aMAA]

Croft, W. (2001) *Radical construction grammar: Syntactic theory in typological perspective.* Oxford University Press. [rMAA]

Crow, T. J. (2002a) Sexual selection, timing and an X-Y homologous gene: Did *Homo sapiens* speciate on the Y chromosome? In: *The speciation of modern Homo sapiens*, ed. T. J. Crow, pp. 197–216. Oxford University Press. [aMAA]

(2002b) *The speciation of modern Homo sapiens.* Oxford University Press. [HF]

Culham, J. C., Danckert, S. L., DeSouza, J. F. X., Gati, J. S., Menon, R. S. & Goodale, M. A. (2003) Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Experimental Brain Research* 153:180–89. [aMAA]

Damasio, A. (1999) *The feeling of what happens: Body and emotion in the making of consciousness.* Harcourt Brace. [HF]

D'Andrade, R. D. (1987) A folk model of the mind. In: *Cultural models in language and thought,* ed. D. Holland & N. Quinn, pp. 112–48. Cambridge University Press. [HF]

Darwin, C. (1871) *The descent of man and selection in relation to sex.* John Murray. [WTF]

(1872/1965) *The expression of the emotions in man and animals*. University of Chicago Press. [aMAA]

Dave, A. S. & Margoliash, D. (2000) Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* 290:812–16. [IMP]

Dawkins, R. (1996) *Climbing Mount Improbable.* Norton. [CB]

Deacon, T. W. (1989) The neural circuitry underlying primate calls and human language. *Human Evolution* 4:367–401. [EG]

(1997) *The symbolic species: The co-evolution of language and the brain.* W. W. Norton. [aMAA, DB, HF]

Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F. & Fazio, F. (1997) Brain activity during observation of actions: Influence of action content and subject's strategy. *Brain* 120:1763–77. [aMAA]

Demiris, Y. & Hayes, G. (2002) Imitation as a dual-route process featuring predictive and learning components: A biologically plausible computational model. In: *Imitation in animals and artifacts,* ed. K. Dautenhahn & C. Nehaniv. MIT Press. [EB]

Demiris, Y. & Johnson, M. (2003) Distributed, predictive perception of actions: A biologically inspired robotics architecture for imitation and learning. *Connection Science Journal* 15(4):231–43. [EB]

Di Ferdinando, A. & Parisi, D. (2004) Internal representations of sensory input reflect the motor output with which organisms respond to the input. In: *Seeing and thinking,* ed. A. Carsetti, pp. 115–41. Kluwer. [DP]

Dixon, R. M. W. (1997) *The rise and fall of languages*. Cambridge University Press. [arMAA]

Dominey, P. F., Arbib, M. A. & Joseph, J.-P. (1995) A model of corticostriatal plasticity for learning associations and sequences, *Journal of Cognitive Neuroscience* 7:311–36. [aMAA]

Dominey, P. F., Hoen, M., Lelekov, T. & Blanc, J. M. (2003) Neurological basis of language and sequential cognition: Evidence from simulation, aphasia and ERP studies. *Brain and Language* 86:207–25. [rMAA]

Donald, M. (1993) *Origins of the modern mind.* Harvard University Press. [WTF]

(1998) Mimesis and the executive suite: Missing links in language evolution. In: *Approaches to the evolution of language: Social and cognitive bases*, ed. J. R. Hurford, M. Studdert-Kennedy & C. Knight, pp. 44–67. Cambridge University Press. [aMAA]

Doupe, A. & Kuhl, P. (1999) Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience* 22:567–631. [arMAA]

Ehrsson, H. H., Fagergren, A., Johansson, R. S. & Forssberg, H. (2003) Evidence for the involvement of the posterior parietal cortex in coordination of fingertip forces for grasp stability in manipulation. *Journal of Neurophysiology* 90:2978–86. [aMAA]

Ehrsson, H. H., Spence, C. & Passingham, R. E. (2004) That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science* 305:875–77. [BI]

Ellis, R. & Tucker, M. (2000) Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology* 91:451–71. [DP]

Emmorey, K. (2002) *Language, cognition, and the brain: Insights from sign language research.* Erlbaum. [arMAA, KE]

Evans, N. (2003) Culture, context and structuration in Australian languages. *Annual Review of Anthropology* 32:13–40. [aMAA]

Fabrega, H. (1997) *Evolution of sickness and healing.* University of California Press. [HF]

(2002) *Origins of psychopathology: Phylogenetic and cultural basis of mental illness.* Rutgers University Press. [HF]

(2004) Consciousness and emotions are minimized. *Behavioral and Brain Sciences* 27(6):736–37. [commentary on Atran, S. & Norenzayan, A. (2004) Religion's evolutionary landscape: Counterintuition, commitment, compassion, communion. *Behavioral and Brain Sciences* 27(6):713–30. [HF]

Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. (2002) Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience* 15:399–402. [aMAA]

Fagg, A. H. & Arbib, M. A. (1992) A model of primate visual-motor conditional learning. *Adaptive Behavior* 1:1–37. [rMAA]

(1998) Modeling parietal-premotor interactions in primate control of grasping. *Neural Networks* 11:1277–303. [aMAA]

Fairman, T. (2000) English pauper letters 1800–34, and the English language. In: *Letter writing as a social practice,* ed. D. Barton & N. Hall, pp. 63–82. John Benjamins. [AW]

Falk, D. (2004a) Prelinguistic evolution in early hominins: Whence motherese. *Behavioral and Brain Sciences* 27(4):491–503. [rMAA, EG, PFM]

(2004b) The roles of infant crying and motherese during prelinguistic evolution in early hominins. *American Journal of Physical Anthropology* (Suppl.) 38:93. [EG]

Ferguson, C. A. (1964) Baby talk in six languages. *American Anthropologist* 66:103–14. [PFM]

Ferrari, P. F., Gallese, V., Rizzolatti, G. & Fogassi, L. (2003) Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience* 17(8):1703–14. [arMAA]

Fichtel, C. & Hammerschmidt, K. (2002) Responses of red-fronted lemurs to experimentally-modified alarm calls: Evidence for urgency-based changes in call structure. *Ethology* 108:763–77. [RMS]

Fischer, J. (1998) Barbary macaques categorize shrill barks into two calls types. *Animal Behaviour* 55:799–807. [RMS]

Fischer, J. & Hammerschmidt, K. (2001) Functional referents and acoustic similarity revisited: The case of Barbary macaque alarm calls. *Animal Cognition* 4:29–35. [RMS]

Fischer, J., Hammerschmidt, K., Seyfarth, R. M. & Cheney, D. L. (2001a) Acoustic features of female chacma baboon barks. *Ethology* 107:33–54. [RMS]

Fischer, J., Metz, M., Cheney, D. L. & Seyfarth, R. M. (2001b) Baboon responses to graded bark variants. *Animal Behaviour* 61:925–31. [RMS]

Floel, A., Ellger, T., Breitenstein, C. & Knecht, S. (2003) Language perception activates the hand motor cortex: Implications for motor theories of speech perception. *European Journal of Neuroscience* 18(3):704–708. [arMAA]

Floreano, D. & Urzelai, J. (2000) Evolutionary robots with on-line self-organization and behavioral fitness. *Neural Networks* 13:431–43. [EB]

Fodor, J. A. (1995) *The language of thought.* Crowell. [rMAA]

Fogassi, L. & Ferrari, P. F. (in press) Mirror neurons, gestures and language evolution. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*. [arMAA]

Fogassi, L., Gallese, V., Fadiga, L. & Rizzolatti, G. (1998) Neurons responding to the sight of goal directed hand/arm actions in the parietal area PF (7b) of the macaque monkey. *Social Neuroscience Abstracts* 24:257. [aMAA]

Foley, R. A. (2001) Evolutionary perspectives on the origins of human social institutions. In: *The origin of human social institutions,* ed. W. G. Runciman, pp. 177–96. Oxford University Press. [HF]

Foundas, A. L., Eure, K. F., Luevano, L. F. & Weinberger, D. R. (1998) MRI asymmetries of Broca's area: The pars triangularis and pars opercularis. *Brain and Language* 64(3):282–96. [AMB]

Friederici, A. D., Kotz, S. A., Steinhauer, K. & von Cramon, D. Y. (2003) The neural basis of the prosody-syntax interplay: The role of the corpus callosum. *Brain and Language* 87:133–34. [BK]

Fritz, J. B., Becker, D., Mishkin, M. & Saunders, R. C. (1999) A comparison of the effects of medial temporal and rhinal cortical lesions on auditory recognition memory in the rhesus monkey. *Society for Neuroscience Abstracts* 25:789. [BH]

Fuster, J. M. (2002) Physiology of executive functions: The perception-action cycle. In: *Principles of frontal lobe function,* ed. D. T. Stuss & R. T. Knight, pp. 96–108. Oxford University Press. [HF]

Galaburda, A. M. & Sanides, F. (1980) Cytoarchitectonic organization of the human auditory cortex. *Journal of Comparative Neurology* 190:597–610. [aMAA, JPR]

Gallese, V. (2003) The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology* 36:171–80. [rMAA]

Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. (1996) Action recognition in the premotor cortex. *Brain* 119:593–609. [aMAA]

Gallese, V. & Goldman, A. (1998) Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2:493–501. [JHGW]

Gamble, C. (1994) *Timewalkers: The prehistory of global colonization.* Harvard University Press. [aMAA]

Geertz, C. (1973) The growth of culture and the evolution of mind. In: *The interpretation of cultures,* ed. C. Geertz, pp. 55–83. Basic Books. [HF]

Gelfand, J. R. & Bookheimer, S. Y. (2003) Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron* 38(5):831–42. [aMAA]

Gemba, H., Miki, N. & Sasaki, K. (1997) Cortical field potentials preceding vocalization in monkeys. *Acta Otolaryngologica Supplemental* 532:96–98. [CB]

Gentilucci, M. (2003a) Grasp observation influences speech production. *European Journal of Neuroscience* 17(1):179–84. [arMAA]

(2003b) Object motor representation and language. *Experimental Brain Research* 153:260–65. [rMAA, DP]

Gerlach, C., Law, I. & Paulson, O. B. (2002) When action turns into words. Activation of motor-based knowledge during categorization of manipulable objects. *Journal of Cognitive Neuroscience* 14(8):1230–39. [aMAA]

Ghazanfar, A. & Hauser, M. D. (2001) The auditory behavior of primates: A neuroethological perspective. *Current Opinions in Neurobiology* 11:712–20. [RMS]

Ghazanfar, A. A., ed. (2003) *Primate audition: Ethology and neurobiology.* CRC Press. [aMAA]

Gibson, J. J. (1979) *The ecological approach to visual perception*. Houghton Mifflin. [aMAA]

Givon, T. (2004) Teaching grammar to apes. Final Project Report, NIH Grant No. 2P01 HD06016–29 SUB: 0015. [DB]

Glover, S. & Dixon, P. (2002) Semantics affect the planning but not control of grasping. *Experimental Brain Research* 146(3):383–87. [aMAA]

Glover, S., Rosenbaum, D. A., Graham, J. & Dixon, P. (2004) Grasping the meaning of words. *Experimental Brain Research* 154(1):103–108. [aMAA]

Goodale, M. A., Milner, A. D., Jakobson, L. S. & Carey, D. P. (1991) A neurological dissociation between perceiving objects and grasping them. *Nature* 349:154–56. [aMAA]

Goodson, J. L. & Bass, A. H. (2002) Vocal-acoustic circuitry and descending vocal pathways in teleost fish: Convergence with terrestrial vertebrates reveals conserved traits. *Journal of Comparative Neurology* 448:298–322. [EG]

Grace, G. (2002) Collateral damage from linguistics? 3: The role of culture-centrism. *Ethnolinguistic Notes* 4(23). Available at: http://www2.hawaii.edu/~grace/elniv23.html. (Online Journal). [AW]

Grafton, S. T., Arbib, M. A., Fadiga, L. & Rizzolatti, G. (1996) Localization of grasp representations in humans by PET: 2. Observation compared with imagination. *Experimental Brain Research* 112:103–11. [aMAA]

Greenfield, P. M. (1991) Language, tools and brain: The ontogeny and phylogeny of hierarchically organized behavior. *Behavioral and Brain Sciences* 14:531–95. [HF, IMP]

Grezes, J. & Decety, J. (2001) Functional anatomy of execution, mental simulation, observation and verb generation of actions: A meta-analysis. *Human Brain Mapping* 12:1–19. [aMAA]

(2002) Does visual perception of object afford action? Evidence from a neuroimaging study. *Neuropsychologia* 40:212–22. [aMAA]

Grezes, J., Armony, J. L., Rowe, J. & Passingham, R. E. (2003) Activations related to "mirror" and "canonical" neurones in the human brain: An fMRI study. *Neuroimage* 18(4):928–37. [aMAA]

Grezes, J., Costes, N. & Decety, J. (1998) Top-down effect of strategy on the perception of human biological motion: A PET investigation. *Cognitive Neuropsychology* 15:553–82. [aMAA]

Guenther, F. H. (1995) Speech sound acquisition, coarticulation and rate effects in a neural network model of speech production. *Psychological Review* 102:594–621. [BH]

Guenther, F. H. & Ghosh, S. S. (2003) A model of cortical and cerebellar function in speech. *Proceedings of the XVth International Congress of Phonetic Sciences,* ed. M. J. Sole, D. Recasens & J. Romero. pp. 169–73. Universitat Autonoma de Barcelona. [BH]

Guillaume, P. (1927/1973) First stages of sentence formation in children's speech. In: *Studies in child language development,* ed. C. A. Ferguson & D. I. Slobin, pp. 522–41. Holt, Rinehart & Winston. [AW]

Hallowell, A. I. (1960) Self, society, and culture in phylogenetic perspective. In: *The evolution of man: Mind, culture, and society, vol. 2, Evolution after Darwin,* ed. S. Tax, pp. 309–71. University of Chicago Press. [HF]

Hamzei, F., Rijntjes, M., Dettmers, C., Glauche, V., Weiller, C. & Buchel, C. (2003) The human action recognition system and its relationship to Broca's area: An fMRI study. *Neuroimage* 19(3):637–44. [aMAA]

Hanakawa, T., Honda, M., Sawamoto, N., Okada, T., Yonekura, Y., Fukuyama, H. & Shibasaki, H. (2002) The role of rostral Brodmann area 6 in mental-operation tasks: An integrative neuroimaging approach. *Cerebral Cortex* 12(11):1157–70. [aMAA]

Harris, R. (2002) Why words really do not stay still. *Times Literary Supplement* 26 July, 5182:30. [DM]

Hast, M. H., Fischer, J. M., Wetzel, A. B. & Thompson, V. E. (1974) Cortical motor representation of the laryngeal muscles in *Macaca mulatto. Brain Research* 73:229–40. [CB]

Hauser, M. D. (1996) *The evolution of communication.* MIT Press/Bradford Books. [aMAA, RMS]

(1998) Functional referents and acoustic similarity: Field playback experiments with rhesus monkeys. *Animal Behaviour* 55:1647–58. [RMS]

Hauser, M. D. & Anderson, K. (1994) Left hemisphere dominance for processing vocalizations in adult, but not infant, rhesus monkeys: Field experiments. *Proceedings of the National Academy of Sciences USA* 91:3946–48. [RMS]

Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298:1569–79. [aMAA]

Heffner, H. E. & Heffner, R. S. (1984) Temporal lobe lesions and perception of species-specific vocalizations by macaques. *Science* 226:75–76. [RMS]

Heilman, K. M., Blonder, L. X., Bowers, D. & Crucian, G. P. (2000) Neurological disorders and emotional dysfunction. In: *The neuropsychology of emotion. Series in affective science,* ed. J. C. Borod. Oxford University Press. [AMB]

Heilman, K. M., Coyle, J. M., Gonyea, E. F. & Geschwind, N. (1973) Apraxia and agraphia in a left-hander. *Brain* 96:21–28. [AMB]

Heilman K. M. & Rothi L. J. G. (2003) Apraxia. In: *Clinical Neuropsychology,* 4th edition, ed. K. M. Heilman & E. Valenstein. Oxford University Press. [AMB]

Heine, B. (1997) *Cognitive foundations of grammar.* Oxford University Press. [aMAA]

Heiser, M., Iacoboni, M., Maeda, F., Marcus, J. & Mazziotta, J. C. (2003) The essential role of Broca's area in imitation. *European Journal of Neuroscience* 17:1981–86. [aMAA]

Herbert, M. R., Harris, G. J., Adrien, K. T., Ziegler, D. A., Makris, N., Kennedy, D. N., Lange, N. T., Chabris, C. F., Bakardjiev, A., Hodgson, J., Takeoka, M., Tager-Flusberg, H. & Caviness, V. S. (2002) Abnormal asymmetry in language association cortex in autism. *Annals of Neurology* 52:588–96. [HT]

Hewes, G. W. (1973) Primate communication and the gestural origin of language. *Current Anthropology* 14:5–24. [aMAA, PFM]

(1996) A history of the study of language origins and the gestural primacy hypothesis. In: *Handbook of human symbolic evolution,* ed. A. Lock & C. R. Peters*.* Oxford University Press. [PFM]

Hickok, G., Bellugi, U. & Klima, E. S. (1998) What's right about the neural organization of sign language? A perspective on recent neuroimaging results. *Trends in Cognitive Sciences* 12:465–68. [aMAA]

Hill, E. L. (2001) Non-specific nature of specific language impairment: A review of the literature with regard to concomitant motor impairments. *International Journal of Language and Communication Disorders* 36:149–71. [HT]

Hill, J. C. (1983) A computational model of language acquisition in the two-year-old. *Cognition and Brain Theory* 6:287–317. [rMAA]

Hobson, R. P. & Bishop, M. (2003) The pathogenesis of autism: Insights from congenital blindness. *Philosophical Transactions of the Royal Society of London B* 358:335–44. [HT]

Hockett, C. F. (1959) Animal "languages" and human language. *Human Biology* 31:32–39. [IMP]

(1978) In search of Jove's brow. *American Speech* 53:243–313. [rMAA, PFM]

Hofstadter, D. (1995) *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought.* Basic Books. [BI]

Holland, J. H. (1992) Genetic algorithms. *Scientific American* 267:66–72. [DP]

Holstein, B. I. (1970) *Use of metaphor to induce innovative thinking in fourth grade children.* Ph.D. thesis, School of Education, Boston University, Boston, MA. [BI]

Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O'Doherty, J., Bullock, P. R. & Polkey, C. E. (2003) Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain* 126:1691–712. [BK]

Horwitz, B. (2005) Integrating neuroscientific data across spatiotemporal scales. *Comptes rendus Biologies* 328:109–18. [BH]

Horwitz, B., Amunts, K., Bhattacharyya, R., Patkin, D., Jeffries, K., Zilles, K. & Braun, A. R. (2003) Activation of Broca's area during the production of spoken and signed language: A combined cytoarchitectonic mapping and PET analysis. *Neuropsychologia* 41(14):1868–76. [aMAA]

Horwitz, B. & Tagamets, M.-A. (2003) Synthetic functional brain imaging. In: *The handbook of brain theory and neural networks,* 2nd edition, ed. M. A. Arbib. MIT Press. [rMAA]

Hunt, G. R. & Gray, R. D. (2002) Diversification and cumulative evolution in New Caledonian crow tool manufacture. *Proceedings of the Royal Society London. Series B* 270:867–74. [aMAA]

Hurford, J. R. (2003) The neural basis of predicate-argument structure. *Behavioral and Brain Sciences* 26(3):261–83. [rMAA]

(2004) Language beyond our grasp: What mirror neurons can, and cannot, do for language evolution. In: *Evolution of communication systems: A comparative approach,* ed. K. Oller, U. Griebel & K. Plunkett, pp. 297–313. MIT Press. [arMAA]

Husain, F. T., Tagamets, M.-A., Fromm, S. J., Braun, A. R. & Horwitz, B. (2004) Relating neuronal dynamics for auditory object processing to neuroimaging activity. *NeuroImage* 21:1701–20. [BH]

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M. C., Mazziotta, J. C. & Rizzolatti, G. (2001) Reafferent copies of imitated actions in the right superior temporal cortex. *Proceedings of the National Academy of Science USA* 98:13995–99. [aMAA]

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C. & Rizzolatti, G. (1999) Cortical mechanisms of human imitation. *Science* 286(5449):2526–28. [aMAA, JHGW]

Indurkhya, B. (1992) *Metaphor and cognition.* Kluwer Academic Publishers. [BI]

(1994) The thesis "all knowledge is metaphorical" and meanings of "metaphor." *Metaphor and Symbolic Activity* 9(1):61–73. [BI]

(1998) On creation of features and change of representation. *Journal of Japanese Cognitive Science Society* 5(2):43–56. [BI]

Ingold, T. (1996) Language is the essence of culture. In: *Key debates in anthropology,* ed. T. Ingold, pp. 147–98. Routledge. [HF]

Ito, M. (2005) Bases and implications of learning in the cerebellum – adaptive control and internal model mechanism. *Progress in Brain Research* 148:95–109. [rMAA]

Izumi, A. & Kojima, S. (2004) Matching vocalizations to vocalizing faces in a chimpanzee (*Pan troglodytes*). *Animal Cognition* 7:179–84. [CB]

Jackendoff, R. (1983) *Semantics and cognition.* MIT Press. [HF]

(1999) Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences* 3(7):272–79. [WTF]

(2002) *Foundations of language: Brain, meaning, grammar, evolution.* Oxford University Press. [aMAA, PFM]

Jackson, J. H. (1878–79) On affections of speech from the disease of the brain. *Brain* 1:304–30; 2:203–22, 323–56. [rMAA]

Janis, I. L. (1982) *Groupthink: Psychological studies of policy decisions and fiascoes,* 2nd edition. Houghton Mifflin. [BK]

Janzen, T. & Shaffer, B. (2002) Gesture as the substrate in the process of ASL grammaticalization. In: *Modality and structure in signed and spoken languages,* ed. R. P. Meier, K. Cormier & D. Quinto-Pozos, pp. 199–223. Cambridge University Press. [KE]

Jarvis, E. D. & Mello, C. V. (2000) Molecular mapping of brain areas involved in parrot vocal communication. *Journal of Comparative Neurology* 419:1–31. [IMP]

Jeannerod, M., Decety, J. & Michel, F. (1994) Impairment of grasping movements. *Neuropsychologia* 32(4):369–80. [aMAA]

Jordan, M. I. & Rumelhart, D. E. (1992) Forward models: Supervised learning with a distal teacher. *Cognitive Science* 16:307–54. [rMAA]

Jürgens, U. (1997) Primate communication: Signaling, vocalization. In: *Encyclopedia of neuroscience,* 2nd edition. Elsevier. [aMAA]

(1998) Neuronal control of mammalian vocalization, with special reference to the squirrel monkey. *Naturwissenschaften* 85(8):376–88. [rMAA, WTF, EG]

(2002) Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews* 26:235–58. [EG]

(2003) From mouth to mouth and hand to hand: On language evolution. *Behavioral and Brain Sciences* 26(2):229–30. [CB]

Jürgens, U. & Alipour, M. (2002) A comparative study on the cortico-hypoglossal connections in primates, using biotine dextranamine. *Neuroscience Letters* 328:245–48. [CB]

Kaas, J. H. (1993) Evolution of multiple areas and modules within neocortex. *Perspectives on Developmental Neurobiology* 1:101–107. [rMAA]

Kanter, S. L., Day, A. L., Heilman, K. M. & Gonzalez Rothi, L. J. (1986) Pure word deafness: A possible explanation of transient deterioration following EC-IC bypass. *Neurosurgery* 18:186–89. [AMB]

Kaplan, J. T. & Iacoboni, M. (2005*)* Brain activations related to hearing and seeing an action. Presented at the Cognitive Neuroscience Society Annual Meeting, New York, April 2005. [JTK]

(in preparation) Multisensory coding of perceived actions in the motor system. [rMAA]

Kegl, J., Senghas, A. & Coppola, M. (1999) Creation through contact: Sign language emergence and sign language change in Nicaragua. In *Language creation and language change: Creolization, diachrony, and development,* ed. M. DeGraff, pp. 179–237. MIT Press. [KE]

Kemmerer, D., Chandrasekaran, B. & Tranel, D. (submitted) A case of impaired verbalization but preserved gesticulation of motion events: Investigating the language-specific representation of space from a neuropsychological perspective. [rMAA]

Kendon, A. (1988) *Sign languages of aboriginal Australia: Cultural, semiotic and communicative perspectives.* Cambridge University Press. [rMAA, DM]

Kertesz, A., Ferro, J. M. & Shewan, C. M. (1984) Apraxia and aphasia: The functional-anatomical basis for their dissociation. *Neurology* 34:40–47. [AMB]

Keysers, C., Kohler, E., Umilta, M. A., Nanetti, L., Fogassi, L. & Gallese, V. (2003) Audiovisual mirror neurons and action recognition. *Experimental Brain Research* 153(4):628–36. [JTK]

Kimura, D. (1993) *Neuromotor mechanisms in human communication* (Oxford Psychology Series No. 20). Oxford University Press/Clarendon Press. [aMAA]

Kirby, S. (2000) *Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners.* In: *The evolutionary emergence of language*, ed. C. Knight, M. Studdert-Kennedy & J. R. Hurford. Cambridge University Press. [arMAA]

Kirzinger, A. & Jürgens, U. (1982) Cortical lesion effects and vocalization in the squirrel monkey. *Brain Research* 233:299–315. [CB]

Kita, S. & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language* 48(1):16–32. [rMAA]

Knight, C. (1991) *Blood relations: Menstruation and the origins of culture*. Yale University Press. [HF]

Kohler, E., Keysers, C., Umiltá, M. A., Fogassi, L., Gallese, V. & Rizzolatti, G. (2002) Hearing sounds, understanding actions: Action representation in mirror neurons. *Science* 297(5582):846–48. [arMAA, JTK, HT]

Kohler, W. (1917/1927) *The mentality of apes*, trans. E. Winter. Routledge & Kegan Paul. (Original work published 1917). [PNP]

Koski, L., Wohlschlager, A., Bekkering, H., Woods, R. P., Dubeau, M. C., Mazziotta, J. C. & Iacoboni, M. (2002) Modulation of motor and premotor activity during imitation of target-directed actions. *Cerebral Cortex* 12:847–55. [aMAA]

Kozima, H., Nakagawa, C. & Yano, H. (2003) Can a robot empathize with people? *Proceedings of the Eighth International Symposium on Artificial Life and Robotics (AROB 08)*, pp. 518–19. [BI]

Kripke, S. A. (1982) *Wittgenstein on rules and private language.* Harvard University Press. [rMAA]

Krubitzer, L. (1998) Constructing the neocortex: Influences on the pattern of organization in mammals. In: *Brain and mind: Evolutionary perspectives,* ed. M. S. Gazzaniga & J. S. Altman, pp. 19–34. Human Frontier Science Program. [rMAA]

Kubovy, M. & Van Valkenburg, D. (2001) Auditory and visual objects. *Cognition* 80:97–126. [BH]

Kuczaj, S. A., Gory, J. D. & Xitco, M. J. (1998) Using programs to solve problems: Imitation versus insight. *Behavioral and Brain Sciences* 21:695–96. [BI]

Kurthen, M. (1992) *Neurosemantik. Grundlagen einer praxiologischen kognitiven Neurowissenschaft.* Ferdinand Enke. [rMAA]

Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413:519–23. [HT]

Langacker, R. (1987) *Foundations of cognitive grammar, vol. 1.* Stanford University Press. [aMAA]

(1991) *Foundations of cognitive grammar, vol. 2.* Stanford University Press. [aMAA]

Lass, R. (1997) *Historical linguistics and language change.* Cambridge University Press. [rMAA]

Leavens, D. A. (2003) Integration of visual and vocal communication: Evidence for Miocene origins. *Behavioral and Brain Sciences* 26:232–33. [CB]

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychological Review* 74(6):431–61. [JTK, JPR]

Lieberman, P. (2002) On the nature and evolution of the neural bases of human language. *Yearbook of Physical Anthropology* 45:36–62. [aMAA]

Liepmann, H. (1905) Die linke Hemisphäre und das Handeln. *Münchener Medizinische Wochenshrift* 49:2322–26, 2375–78. [AMB]

(1920) Apraxia. *Ergebnisse der Gesamten Medizin* 1:516–43. [AMB]

Linnankoski, I., Laakso, M.-L., Aulanko, R. & Leinonen, L. (1994) Recognition of emotions in macaque vocalizations by children and adults. *Language & Communication* 14:183–92. [BK]

Locke, J. L. (1993) *The child's path to spoken language.* Harvard University Press. [PFM]

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., Pickles, A. & Rutter, M. (2000) The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders* 30:205–23. [JHGW]

MacNeilage, P. F. (1998) The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21:499–546. [aMAA, HF, PFM]

(2003) Mouth to hand and back again? Could language have made those journeys? *Behavioral and Brain Sciences* 26(2):233–34. [rMAA, PFM]

MacNeilage, P. F. & Davis, B. L. (1990) Acquisition of speech: Frames, then content. In: *Attention and Performance XIII*, ed. M Jeannerod. Erlbaum. [PFM]

(2000) Origin of the internal structure of word forms. *Science* 288:527–31. [PFM]

(2004) Baby talk and the origin of the word. Paper presented at the Fifth International Conference on the Evolution of Language, Leipzig, Germany, April 2004. [PFM]

(in press a) Baby talk and the emergence of first words. *Behavioral and Brain Sciences*. [PFM]

(in press b) The frame/content theory of evolution of speech: Comparison with a gestural origins theory. *Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems.* [arMAA, PFM]

Macphail, E. M. (2000) The search for a mental Rubicon. In: *The evolution of cognition,* ed. C. Heyes & L. Huber, pp. 253–72. MIT Press. [rMAA, HF]

Makino, T. & Aihara, K. (2003) Self-observation principle for estimating the other's internal state: A new computational theory of communication. *Mathematical Engineering Technical Reports* METR 2003–36, Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo. [rMAA]

Manthey, S., Schubotz, R. I. & von Cramon, D. Y. (2003) Premotor cortex in observing erroneous action: An fMRI study. *Cognitive Brain Research* 15(3):296–307. [aMAA]

Marino, L. (2002) Convergence of complex cognitive abilities in cetaceans and primates. *Brain, Behavior and Evolution* 59:21–32. [EG]

Marom, Y., Maistros, G. & Hayes, G. (2002) Toward a mirror system for the development of socially-mediated skills. In: *Proceedings of the Second International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Edinburgh, Scotland,* vol. 94, ed. C. Prince, Y. Demiris, Y. Marom, H. Kozima & C. Balkenius. Lund University Cognitive Studies. [EB]

Matelli, M., Camarda, R., Glickstein, M. & Rizzolatti, G. (1986) Afferent and efferent projections of the inferior area 6 in the macaque monkey. *Journal of Comparative Neurology* 251:281–98. [aMAA]

McNeill, D. (1992) *Hand and mind: What gestures reveal about thought.* University of Chicago Press. [DM]

(2000) Introduction. In: *Language and gesture,* ed. D. McNeill, pp. 1–10. Cambridge University Press. [DM]

Meltzoff, A. N. & Decety, J. (2003) What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358(1431):491–500. [rMAA, JHGW]

Meltzoff, A. N. & Prinz, W. (2002) The imitative mind: Development, evolution and brain bases. In: *Cambridge studies in cognitive development,* ed. K. W. Fischer & G. Hatano. Cambridge University Press. [JHGW]

Miall, R. C. (2003) Connecting mirror neurons and forward models. *NeuroReport* 14:2135–37. [rMAA]

Miller, G. F. (2001) *The mating mind: How sexual choice shaped the evolution of human nature.* Anchor Books. [CY]

Minter, M. E., Hobson, R. P. & Bishop, M. (1998) Congenital visual impairment and "theory of mind." *British Journal of Developmental Psychology* 16:183–96. [HT]

Mitz, A. R., Godshalk, M. & Wise, S. P. (1991) Learning-dependent neuronal activity in the premotor cortex: Activity during the acquisition of conditional motor associations. *Journal of Neuroscience* 11(6):1855–72. [rMAA]

Moore, B. R. (1992) Avian movement imitation and a new form of mimicry: Tracing the evolution of a complex form of learning. *Behaviour* 12:231–63. [WTF, BI]

Murdock, G. P. (1959) Cross-language parallels in parental kin terms. *Anthropological Linguistics* 1:1–5. [rMAA, PFM]

Myowa-Yamakoshi, M. & Matsuzawa, T. (1999) Factors influencing imitation of manipulatory actions in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology* 113:128–36. [aMAA]

Nagy, E. & Molnar, P. (2004) Homo imitans or homo provocans? The phenomenon of neonatal initiation. *Infant Behavior and Development* 27:57–63. [rMAA]

Newbury, D. F., Bonora, E., Lamb, J. A., Fisher, S. E., Lai, C. S., Baird, G., Jannoun, L., Slonims, V., Stott, C. M., Merricks, M. J., Bolton, P. F., Bailey, A. J., Monaco, A. P., (International Molecular Genetic Study of Autism Consortium) (2002) FOXP2 is not a major susceptibility gene for autism or specific language impairment. *American Journal of Human Genetics* 70:1318–27. [HT]

Nishitani, N., Avikainen, S. & Hari, R. (2004) Abnormal imitation-related cortical activation sequences in Asperger's syndrome. *Annals of Neurology* 55:558–62. [HT]

Noble, W. & Davidson, I. (1996) *Human evolution, language, and mind: A psychological and archeological inquiry.* Cambridge University Press. [HF]

Nowak, M. A., Plotkin, J. B. & Jansen, V. A. A. (2000) The evolution of syntactic communication. *Nature* 404:495–98. [aMAA]

Ochipa, C., Rothi, L. J. G. & Heilman, K. M. (1994) Conduction apraxia. *Journal of Neurology, Neurosurgery and Psychiatry* 57:1241–44. [AMB]

Owren, M. J., Hopp, S. L. & Seyfarth, R. M. (1992) Categorical vocal signaling in nonhuman primates. In: *Nonverbal vocal communication: Comparative and developmental approaches,* ed. H. Papousek, U. Jürgens & M. Papousek. Cambridge University Press. [RMS]

Owren, M. J., Seyfarth, R. M. & Cheney, D. L. (1997) The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cynocephalus ursinus*). *Journal of the Acoustical Society of America* 101:2951–63. [RMS]

Oztop, E. & Arbib, M. A. (2002) Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics* 87:116–40. [aMAA, EB]

Oztop, E., Bradley, N. S. & Arbib, M. A. (2004) Infant grasp learning: A computational model. *Experimental Brain Research* 158:480–503. [aMAA]

Papagno, C., della Sala, S. & Basso, A. (1993) Ideomotor apraxia without aphasia and aphasia without apraxia: The anatomical support for a double dissociation. *Journal of Neurology, Neurosurgery and Psychiatry* 56:286–89. [AMB]

Passingham, R. (1993) *The frontal lobes and voluntary action.* Oxford University Press. [aMAA]

Patterson, D. K. & Pepperberg, I. M. (1994) A comparative study of human and parrot phonation: Acoustic and articulatory correlates of vowels. *Journal of the Acoustical Society of America* 96:634–48. [IMP]

(1998) Acoustic and articulatory correlates of stop consonants in a parrot and a human subject. *Journal of the Acoustical Society of America* 106:491–505. [IMP]

Paulesu, E., Goldacre, B., Scifo, P., Cappa, S. F., Gilardi, M. C., Castigllioni, I., Perani, D. & Fazio, F. (1997) Functional heterogeneity of left inferior frontal cortex as revealed by MRI. *Neuroreport* 8:2011–17. [AMB]

Payne, K. & Payne, R. (1985) Large scale changes over 19 years in songs of humpback whales in Bermuda. *Zeitschrift für Tierpsychologie* 68:89–114. [EG]

Peirce, A. (1985) A review of attempts to condition operantly alloprimate vocalizations. *Primates* 26:202–13. [RMS]

Pepperberg, I. M. (1992) Proficient performance of a conjunctive, recursive task by an African Grey parrot (*Psittacus erithacus*). *Journal of Comparative Psychology* 106:295–305. [IMP]

(1999) *The Alex studies: Cognitive and communicative abilities of Grey parrots.* Harvard University Press. [IMP]

(2005a) Grey parrots do not always "parrot": Phonological awareness and creation of new labels from existing vocalizations. In: *Proceedings of the Third International Symposium on Imitation in Animals and Artifacts, 2005 Artificial Intelligence and the Simulation of Behavior Convention.* pp. 97–104. [IMP]

(2005b) Insights into imitation in Grey parrots (*Psittacus erithacus*). In: *Perspectives on imitation: From neuroscience to social science,* ed. S. L. Hurley & N. Chader, pp. 243–62. MIT Press. [IMP]

Pepperberg, I. M., Brese, K. J. & Harris, B. J. (1991) Solitary sound play during acquisition of English vocalizations by an African Grey Parrot (*Psittacus erithacus*): Possible parallels with children's monologue speech. *Applied Psycholinguistics* 12:151–78. [IMP]

Pepperberg, I. M. & Brezinsky, M. V. (1991) Acquisition of a relative class concept by an African Grey Parrot (*Psittacus erithacus*): Discriminations based on relative size. *Journal of Comparative Psychology* 105:286–94. [IMP]

Pepperberg, I. M. & Gordon, J. D. (2005) Number comprehension by an African grey parrot (*Psittacus erithacus*), including a zero-like concept. *Journal of Comparative Psychology* 119:197–209. [IMP]

Pepperberg, I. M. & Shive, H. A. (2001) Simultaneous development of vocal and physical object combinations by a Grey Parrot (*Psittacus erithacus*): Bottle caps, lids, and labels. *Journal of Comparative Psychology* 115:376–84. [IMP]

Perrett, D. I., Mistlin, A. J., Harries, M. H. & Chitty, A. J. (1990) Understanding the visual appearance and consequence of hand actions. In: *Vision and action: The control of grasping,* ed. M. A. Goodale, pp. 163–342. Ablex. [aMAA]

Perrett, D., Rolls, E. T. & Cann, W. (1982) Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research* 47:329–42. [DB]

Perrett, D., Smith, P. A. J., Mistlin, A. J., Chitty, A. J, Head, A. S., Potter, D. D, Broenniman, R., Milner, A. P. & Jeeves, M. A. (1985) Visual analysis of body movements by neurones in the temporal cortex of the macaque monkey. *Behavior and Brain Research* 16(2–3):153–70. [DB]

Peters, A. M. (1983) *Units of language acquisition.* Cambridge University Press. [rMAA, AW]

Petersen, M. R., Beecher, M. D., Zoloth, S. L., Moody, D. B. & Stebbins, W. C. (1978) Neural lateralization of species-specific vocalizations by Japanese macaques. *Science* 202:324–27. [RMS]

Petitto, L. A. & Marentette, P. F. (1991) Babbling in the manual mode: Evidence for the ontogeny of language. *Science* 251:1493–96. [BH]

Petrides, M. & Pandya, D. N. (1984) Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *Journal of Comparative Neurology* 228:105–16. [aMAA]

Pfeiffer, J. E. (1982) *The creative explosion. An inquiry into the origins of art and religion.* Cornell University Press. [HF]

Pinker, S. (1995) *The language instinct: How the mind creates language.* Harper Perennial. [CB]

Ploog, D. (1988) Neurobiology and pathology of subhuman vocal communication and human speech. In: *Primate vocal communication,* ed. D. Todt, P. Goedeking & D. Symmes. Springer Verlag. [EG]

Plumer, T. K. & Striedter, G. F. (1997) Auditory and vocalization related activity in the vocal control system of budgerigars. *Society for Neuroscience Abstracts* 100:10. [IMP]

(2000) Auditory responses in the vocal motor system of budgerigars. *Journal of Neurobiology* 42:79–94. [IMP]

Podell, K., Wisniewski, K. & Lovell, M. R. (2001) The assessment of echopraxia as a component of executive control deficit in traumatic brain injury. *Brain and Cognition* 47(1–2):349–53. [rMAA]

Poremba, A., Malloy, M., Saunders, R. C., Carson, R. E., Herscovitch, P. & Mishkin, M. (2004) Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature* 427:448–51. [RMS]

Prell, C. G., Hauser, M. D. & Moody, D. B. (2002) Discrete or graded variation within rhesus monkey screams? Psychophysical experiment on classification. *Animal Behaviour* 63:47–62. [RMS]

Provine, R. R. (1986) Yawning as a stereotyped action pattern and releasing stimulus. *Ethology* 72:109–22. [RRP]

(1989) Faces as releasers of contagious yawning: An approach to face detection using normal human subjects. *Bulletin of the Psychonomic Society* 27:211–14. [RRP]

(1993) Laughter punctuates speech: Linguistic, social and gender contexts of laughter. *Ethology* 95:291–98. [RRP]

(1996) Laughter. *American Scientist* 84:38–45. [RRP]

(2000) *Laughter: A scientific investigation*. Viking. [RRP]

Prudkov, P. N. (1999a) Human evolution expanded brains to increase motivational ability, not expertise capacity. *Psycoloquy* 10(012). Available at: http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?10.012. (Online Journal). [PNP]

(1999b) Origin of culture: Evolution applied another mechanism. *Psycoloquy* 10 (037). Available at: http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?10.037. (Online Journal.) [PNP]

(1999c) The role of motivation in the origin of language. *Psycoloquy* 10(069). Available at: http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?10.069. (Online Journal.) [PNP]

Prudkov, P. N. & Rodina, O. N. (1999) Synthesis of purposeful processes. *Psycoloquy* 10(070). Available at: http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?10.070. (Online Journal.) [PNP]

Pulvermüller, F. (2002) A brain perspective on language mechanisms: From discrete neuronal ensembles to serial order. *Progress in Neurobiology* 67(2):85–111. [rMAA]

Quine, W. V. (1960) *Word and object.* MIT Press. [CY]

Rauschecker, J. P. (1998) Cortical processing of complex sounds. *Current Opinion in Neurobiology* 8:516–21. [JPR]

Rauschecker, J. P. & Tian, B. (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America* 97:11800–806. [JPR]

Raymer, A. M. & Ochipa, C. (1997) Conceptual praxis. In: *Apraxia: The neuropsychology of action,* ed. L. J. G. Rothi & K. M. Heilman. Psychology Press. [AMB]

Reiss, D. & McCowan, B. (1993) Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): Evidence for vocal learning. *Journal of Comparative Psychology* 107:301–12. [WTF]

Rendall, D. (2003) Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *Journal of the Acoustical Society of America* 113:3390–402. [RMS]

Rendall, D., Seyfarth, R. M., Cheney, D. L. & Owren, M. J. (1999) The meaning and function of grunt variants in baboons. *Animal Behaviour* 57:583–92. [RMS]

Ribeiro, S., Cecchi, G. A., Magnasco, M. O. & Mello, C. (1998) Toward a song code: Evidence for a syllabic representation in the canary brain. *Neuron* 21:359–71. [IMP]

Richards, D. G., Wolz, J. P. & Herman, L. M. (1984) Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus. Journal of Comparative Psychology* 98:10–28. [WTF]

Ritt, N. (2004) *Selfish sounds and linguistic evolution*. Cambridge University Press. [rMAA]

Rizzolatti, G. & Arbib, M. A. (1998) Language within our grasp. *Trends in Neurosciences*. 21(5):188–94. [aMAA, CB, JPR]

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G. & Matelli, M. (1988) Functional organization of inferior area 6 in the macaque monkey II. Area F5 and the control of distal movements. *Experimental Brain Research* 71:491–507. [aMAA]

Rizzolatti, G. & Craighero, L. (2004) The mirror-neuron system. *Annual Review of Neuroscience* 27:169–92. [HT]

Rizzolatti, G., Fadiga L., Gallese, V. & Fogassi, L. (1996a) Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* 3:131–41. [aMAA]

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Perani, D. & Fazio, F. (1996b) Localization of grasp representations in humans by positron emission tomography: 1. Observation versus execution. *Experimental Brain Research* 111:246–52. [aMAA]

Rizzolatti, R., Fogassi, L. & Gallese, V. (2001) Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience* 2:661–70. [arMAA]

Rizzolatti G. & Luppino G. (2001) The cortical motor system. *Neuron* 31:889–901. [aMAA]

(2003) Grasping movements: Visuomotor transformations. In: *The handbook of brain theory and neural networks,* 2nd edition, ed. M. A. Arbib, pp. 501–504. MIT Press. [aMAA]

Rizzolatti, G., Luppino, G. & Matelli M. (1998) The organization of the cortical motor system: New concepts. *Electroencephalography and Clinical Neurophysiology* 106:283–96. [aMAA]

Rolls, E. T. & Arbib, M. A. (2003) Visual scene perception. In: *The handbook of brain theory and neural networks*, 2nd edition, ed. M. A. Arbib. MIT Press. [aMAA]

Romanski, L. M. & Goldman-Rakic, P. S. (2002) An auditory domain in primate prefrontal cortex. *Nature Neuroscience* 5:15–16. [CB, JPR]

Ross, E. D. (1981) The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurol*ogy 38(9):561–69. [AMB, BK]

Rothi, L. J. G., Raade, A. S. & Heilman, K. M. (1994) Localization of lesions in limb and buccofacial apraxia. In: *Localization and neuroimaging in neuropsychology,* ed. A. Kertesz. Academic Press. [AMB]

Ruppin, E. (2002) Evolutionary autonomous agents: A neuroscience perspective. *Nature Reviews Neuroscience* 3(2):132–41. [EB]

Saunders, R. C., Fritz, J. B. & Mishkin, M. (1998) The effects of rhinal cortical lesions on auditory short-term memory (STM) in the rhesus monkey. *Society for Neuroscience Abstracts* 24:1907. [BH]

Scherer, K. R. (1986) Vocal affect expression: A review and a model for future research. *Psychological Bulletin* 99:143–65. [BK]

Schubotz, R. I. & von Cramon, D. Y. (2003) Functional-anatomical concepts of human premotor cortex: Evidence from fMRI and PET studies. *NeuroImage* 20 (Suppl. 1):S120–31. [JTK]

Scott, S. K., Blank, C. C., Rosen, S. & Wise, R. J. (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123 (Pt 12):2400–406. [JPR]

Seddoh, S. A. (2002) How discrete or independent are "affective prosody" and "linguistic prosody"? *Aphasiology* 16(7):683–92. [BK]

Seltzer, B. & Pandya, D. N. (1994) Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: A retrograde tracer study. *Journal of Comparative Neurology* 15:445–63. [aMAA]

Seyfarth, R. M. & Cheney, D. L. (2003a) Meaning and emotion in animal vocalizations. *Annals of the New York Academy of Sciences* 1000:32–55. [CB]

(2003b) Signalers and receivers in animal communication. *Annual Review of Psychology* 54:145–73. [RMS]

Seyfarth, R. M., Cheney, D. L. & Marler, P. (1980) Monkey responses to three different alarm calls: Evidence for predator classification and semantic communication. *Science* 210:801–803. [RMS]

Shallice, T. (1988) *From neuropsychology to mental structure.* Cambridge University Press. [HF]

Shaw, R. & Turvey, M. (1981) Coalitions as models of ecosystems: A realist perspective on perceptual organization. In: *Perceptual organization,* ed. M. Kubovy & J. R. Pomerantz, pp. 343–415. Erlbaum. [rMAA]

Simner, M. L. (1971) Newborn's response to the cry of another infant. *Developmental Psychology* 5:136–50. [RRP]

Skelly, M., Schinsky, L., Smith, R. W. & Fust, R. S. (1974) American Indian Sign (Amerind) as a facilitator of verbalization for the oral verbal apraxic. *Journal of Speech and Hearing Disorders* 39:445–56. [AMB]

Smith, E. E. & Jonides, J. (1998) Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Sciences U.S.A.* 95:12061–68. [CB]

Snowdon, C. T. (1989) Vocal communication in New World monkeys. *Journal of Human Evolution* 18:611–33. [EG]

(1990) Language capacities of nonhuman animals. *Yearbook of Physical Anthropology* 33:215–43. [RMS]

Snowdon, C. T., French, J. A. & Cleveland, J. (1986) Ontogeny of primate vocalizations: Models from birdsong and human speech. In: *Current perspectives in primate social behavior,* ed. D. Taub & F. E. King. Van Nostrand Rheinhold. [RMS]

Speedie, L. J., Wertman, E., Tair, J. & Heilman, K. M. (1993) Disruption of automatic speech following a right basal ganglia lesion. *Neurology* 43:1768–74. [AMB]

Stoel-Gammon, C. & Otomo, K. (1986) Babbling development of hearing-impaired and normally hearing subjects. *Journal of Speech and Hearing Disorders* 51:33–41. [BH]

Stokoe, W. C. (2001) *Language in hand: Why sign came before speech.* Gallaudet University Press. [aMAA]

Striedter, G. (1994) The vocal control pathways in budgerigars differ from those in songbirds. *Journal of Comparative Neurology* 343:35–56. [IMP]

(2004) *Principles of brain evolution.* Sinauer Associates. [rMAA]

Studdert-Kennedy, M. (2002) Mirror neurons, vocal imitation and the evolution of particulate speech. In: *Mirror neurons and the evolution of brain and language,* ed. M. Stamenov & V. Gallese, pp. 207–27. John Benjamins. [rMAA]

Studdert-Kennedy, M. G. & Lane, H. (1980) Clues from the differences between signed and spoken languages. In: *Signed and spoken language: Biological constraints on linguistic form,* ed. U. Bellugi & M. Studdert-Kennedy. Verlag Chemie. [PFM]

Suddendorf, T. (1999) The rise of the metamind. In: *The descent of mind: Psychological perspectives on hominid evolution,* ed. M. C. Corballis & S. E. G. Lea, pp. 218–60. Oxford University Press. [HF]

Suddendorf, T. & Corballis, M. C. (1997) Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs* 123(2):133–67. [HF, PNP]

Suddendorf, T. & Whiten, A. (2001) Mental evolution and development: Evidence for secondary representation in children, great apes, and other animals. *Psychological Bulletin* 127:629–50. [JHGW]

Supalla, T. (1986) The classifier system in American Sign Language. In: *Noun classes and categorization, vol. 7 of Typological studies in language,* ed. C. Craig. John Benjamins. [aMAA]

Supalla, T. & Newport, E. (1978) How many seats in a chair? The derivation of nouns and verbs in American sign language. In: *Understanding language through sign language research*, ed. P. Siple. Academic Press. [aMAA]

Sutton, D. & Jürgens, U. (1988) Neural control of vocalization. In: *Comparative primate biology, vol. 4, Neurosciences.* Alan R. Liss. [EG]

Tager-Flusberg, H. (1997) The role of theory of mind in language acquisition: Contributions from the study of autism. In: *Research on communication and language disorders: Contributions to theories of language development,* ed. L. Adamson & M. A. Romski. Paul Brookes. [HT]

Taira, M., Mine, S., Georgopoulos, A. P., Murata, A. & Sakata, H. (1990) Parietal cortex neurons of the monkey related to the visual guidance of hand movement. *Experimental Brain Research* 83:29–36. [aMAA]

Tallerman, M. (2004) Analyzing the analytic: Problems with holistic theories of protolanguage. Paper presented at the Fifth Biennial Conference on the Evolution of Language, Leipzig, Germany, April 2004. [rMAA, DB]

Talmy, L. (2000) *Towards a cognitive semantics,* 2 vols. MIT Press. [aMAA]

Théoret, H., Halligan, E., Kobayashi, M., Fregni, F., Tager-Flusberg, H. & Pascual-Leone, A. (2005) Impaired motor facilitation during action observation in individuals with autism spectrum disorder. *Current Biology* 15(3):R84–85. [HT]

Tian, B., Reser, D., Durham, A., Kustov, A. & Rauschecker, J. P. (2001) Functional specialization in rhesus monkey auditory cortex. *Science* 292:290–93. [JPR]

Tomasello, M. (1999a) *The cultural origins of human cognition.* Harvard University Press. [HF]

(1999b) The human adaptation for culture. *Annual Review of Anthropology* 28:509–29. [aMAA]

Tomasello, M. & Call, J. (1997) *Primate cognition.* Oxford University Press. [aMAA]

Tonkonogy, J. & Goodglass, H. (1981) Language function, foot of the third frontal gyrus, and rolandic operculum. *Archives of Neurology* 38:486–90. [AMB]

Tononi, G. & Edelman, G. M. (1998) Consciousness and complexity. *Science* 282:1846–51. [HF]

Trevarthen, C. (2001) The neurobiology of early communication: Intersubjective regulations in human brain development. In: *Handbook on brain and behavior in human development,* ed. A. F. Kalverboer & A. Gramsbergen, pp. 841–82. Kluwer. [rMAA]

Tsiotas, G., Borghi, A. & Parisi, D. (in press) Objects and affordances: An Artificial Life simulation. *Proceeding of the XVII Annual Meeting of the Cognitive Science Society.* [DP]

Tucker, D. M., Watson, R. T. & Heilman, K. M. (1977) Affective discrimination and evocation in patients with right parietal disease. *Neurology* 17:947–50. [AMB]

Tucker, M. & Ellis, R. (1998) On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance* 24(3):830–46. [DP]

(2004) Action priming by briefly presented objects. *Acta Psychologica* 116:185–203. [DP]

Ullman, M. (2004) Contributions of memory circuits to language: The declarative procedural model. *Cognition* 92:231–70. [rMAA]

Umiltá, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C. & Rizzolatti, G. (2001) I know what you are doing. A neurophysiological study. *Neuron* 31:155–65. [aMAA]

Ungerleider, L. G. & Mishkin, M. (1982) Two cortical visual systems. In: *Analysis of visual behaviour,* ed. D. J. Ingle, M. A. Goodale & R. J. W. Mansfield, pp. 549–86. MIT Press. [JPR]

Vaccari, O. & Vaccari, E. E. (1961) *Pictorial Chinese-Japanese characters*, 4th edition. Charles E. Tuttle. [aMAA]

Valenstein, E. & Heilman, K. M. (1979) Apraxic agraphia with neglect-induced paragraphia. *Archives of Neurology* 36:506–508. [AMB]

Ventner A., Lord, C. & Schopler, E. (1992) A follow-up study of high-functioning autistic children. *Journal of Child Psychology and Psychiatry* 33:489–507. [HT]

Visalberghi, E. & Fragaszy, D. (2002) "Do monkeys ape?" Ten years after. In: *Imitation in animals and artifacts*, ed. C. Nehaniv & K. Dautenhahn. MIT Press [aMAA]

Voelkl, B. & Huber, L. (2000) True imitation in marmosets? *Animal Behaviour* 60:195–20. [aMAA]

Wang, X., Merzenich, M. M., Beitel, R. & Schreiner, C. E. (1995) Representation of species-specific vocalization in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics. *Journal of Neurophysiology* 74:2685–706. [RMS]

Watson, R. T., Fleet, W. S., Gonzalez-Rothi, L. & Heilman, K. M. (1986) Apraxia and the supplementary motor area. *Archives of Neurology* 43:787–92. [AMB]

Weydemeyer, W. (1930) An unusual case of mimicry by a catbird. *Condor* 32:124–25. [WTF]

Wheeler, M. A., Stuss, D. T. & Tulving, E. (1997) Toward a theory of episodic memory: The frontal lobes and autonoetic consciousness. *Psychological Bulletin* 121(3):331–54. [HF]

Whiten A. & Byrne, R. W. (1997) *Machiavellian intelligence II.: Evaluations and extensions.* Cambridge University Press. [JHGW]

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama Y., Tutin, C. E. G., Wrangham, R. W. & Boesch, C. (2001) Charting cultural variation in chimpanzees. *Behaviour* 138:1481–516. [arMAA]

Whiten, A., Horner, V., Litchfield, C. & Marshall-Pescini, S. (2004) How do apes ape? *Learning and Behaviour* 32:36–52. [JHGW]

Wierzbicka, A. (1992) *Semantics, culture, and cognition: Universal human concepts in culture-specific configurations.* Oxford University Press. [HF]

(1993) A conceptual basis for cultural psychology. *Ethos* 21(2):205–31. [HF]

Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B. & Grodd, W. (1996) Functional lateralization of speech production at primary motor cortex: A fMRI study. *NeuroReport* 7:2791–95. [AMB]

Wilkins, W. K. & Wakefield, J. (1995) Brain evolution and neurolinguistic preconditions. *Behavioral and Brain Sciences* 18:161–226. [HF]

Williams, H. (1989) Multiple representations and auditory-motor interactions in the avian song system. *Annals of the New York Academy of Sciences* 563:148–64. [IMP]

Williams, J. H. G., Whiten, A., Suddendorf, T. & Perrett, D. I. (2001) Imitation, mirror neurons and autism. *Neuroscience and Biobehavioral Reviews* 25:287–95. [HT, JHGW]

Wilson, S. M., Saygin, A. P., Sereno, M. I. & Iacoboni, M. (2004) Listening to speech activates motor areas involved in speech production. *Nature Neuroscience* 7(7):701–702. [CB, JTK]

Wittgenstein, L. (1958) *Philosophical investigations.* Blackwell. [rMAA]

Wolpert, D. M., Doya, K. & Kawato, M. (2003) A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 358:593–602. [rMAA]

Worden, R. (1998) The evolution of language from social intelligence. In: *Approaches to the evolution of language,* ed. J. R. Hurford, M. Studdert-Kennedy & C. Knight. Cambridge University Press. [RMS]

Wray, A. (1998) Protolanguage as a holistic system for social interaction. *Language & Communication* 18:47–67. [aMAA, DB, HF, AW]

(2000) Holistic utterances in protolanguage: The link from primates to humans. In: *The evolutionary emergence of language: Social function and the origins of linguistic form,* ed. C. Knight, M. Studdert-Kennedy & J. Hurford, pp. 285–302. Cambridge University Press. [aMAA, DB, HF, AW]

(2002a) *Formulaic language and the lexicon.* Cambridge University Press. [rMAA, WTF, AW]

(2002b) Dual processing in protolanguage: Performance without competence. In: *The transition to language,* ed. A. Wray, pp. 113–37. Oxford University Press. [AW]

(2004) "Here's one I prepared earlier": Formulaic language learning on television. In: *The acquisition and use of formulaic sequences,* ed. N. Schmitt, pp. 249–68. John Benjamins. [AW]

Yu, C. & Ballard, D. H. (2004) A multimodal learning interface for grounding spoken language in sensory perceptions. *ACM Transactions on Applied Perception* 1:57–80. [CY]

Yu, C., Ballard, D. H. & Aslin, R. N. (2003) The role of embodied intention in early lexical acquisition. In*: Proceedings of the 25th Annual Conference of the Cognitive Science Society*, pp. 1293–98, ed. R. Alterman & D. Kirsh. Erlbaum. A long version is submitted to *Journal of Cognitive Science.* [CY]

Zatorre, R. J., Bouffard, M. & Belin, P. (2004) Sensitivity to auditory object features in human temporal neocortex. *Journal of Neuroscience* 24:3637–42. [JPR]

Zentall, T. & Akins, C. (2001) Imitation in animals: Evidence, function, and mechanisms. In: *Avian visual cognition,* ed. R. G. Cook. Available online at: http://www.pigeon.psy.tufts.edu/avc/zentall/. [BI]

Zeshan, U. (2003) Aspects of Turk Isaret Dili (Turkish Sign Language). *Sign Language & Linguistics* 6(1):43–75. [KE]

Zoloth, S. R., Petersen, M. R., Beecher, M. D., Green, S., Marler, P., Moody, D. B. & Stebbins, W. C. (1979) Species-specific perceptual processing of vocal sounds by monkeys. *Science* 204:870–72. [RMS]

Zuberbuhler, K. (2003) Referential signaling in nonhuman primates: Cognitive precursors and limitations for the evolution of language. *Advances in the Study of Behavior* 33:265–307. [RMS]

Zuberbuhler, K., Cheney, D. L. & Seyfarth, R. M. (1999) Conceptual semantics in a nonhuman primate. *Journal of Comparative Psychology* 113:33–42. [RMS]

Zukow-Goldring, P. (1996) Sensitive caregivers foster the comprehension of speech: When gestures speak louder than words. *Early Development and Parenting* 5(4):195–211. [rMAA]