

Toward the Language-Ready Brain: Biological Evolution and Primate Comparisons

Michael A. Arbib¹

Published online: 1 July 2016
© Psychonomic Society, Inc. 2016

Abstract The approach to language evolution suggested here focuses on three questions: How did the human brain evolve so that humans can develop, use, and acquire languages? How can the evolutionary quest be informed by studying brain, behavior, and social interaction in monkeys, apes, and humans? How can computational modeling advance these studies? I hypothesize that the brain is *language ready* in that the earliest humans had protolanguages but not languages (i.e., communication systems endowed with rich and open-ended lexicons and grammars supporting a compositional semantics), and that it took cultural evolution to yield societies (a cultural constructed niche) in which language-ready brains could become language-using brains. The mirror system hypothesis is a well-developed example of this approach, but I offer it here not as a closed theory but as an evolving framework for the development and analysis of conflicting subhypotheses in the hope of their eventual integration. I also stress that computational modeling helps us understand the evolving role of mirror neurons, not in and of themselves, but only in their interaction with systems “beyond the mirror.” Because a theory of evolution needs a clear characterization of what it is that evolved, I also outline ideas for research in neurolinguistics to complement studies of the evolution of the language-ready brain. A clear challenge is to go beyond models of speech comprehension to include sign language and models of production, and to link language to visuomotor interaction with the physical and social world.

Keywords Language evolution · Mirror neurons · Primates · Computational modeling

Computational Comparative Neuroprimatology

Much work on language evolution sidesteps any concern with the brain mechanisms that support language. By contrast, this paper sketches an approach grounded by the question:

Q1. *How did the human brain evolve so that humans can develop, use, and acquire languages?*

Any answer to this question is heavily conditioned on whether or not one accepts the hypothesis that the human brain has innate structures that prespecify the overall structure of human grammars (as in the principles and parameters version of universal grammar). The present approach is framed by the counterhypothesis that the *H. sapiens* brain was *language ready*, but that it took tens of millennia of cultural evolution before human had languages and *language-using brains*, even though the brain genome had not changed significantly in the interim. More specifically, the hypothesis is that biological evolution and cultural evolution had together yielded brains and social structures in early *H. sapiens* that could support rudimentary systems of communication, *protolanguages*, with a somewhat open set of “protowords,” but not *languages* in the sense of communication systems endowed with rich and open-ended lexicons and grammars supporting a compositional semantics. It thus took cultural evolution—*niche construction* (Laland, Odling-Smee, & Feldman, 2000)—to yield societies in which *language-ready* brains could become *language-using* brains.

How, then, does one learn about the evolutionary history of human brains when brains do not fossilize? Some clues can be gleaned from the study of endocasts, the indentation in

✉ Michael A. Arbib
arbib@usc.edu

¹ Computer Science, Neuroscience and the ABLE Project (Action, Brain, Language, Evolution), University of Southern California, Los Angeles, CA 90089, USA

fossilized skulls of *Homo* and *Australopithecus* that offer blurred impressions of the long gone cerebral cortex, but the approach here is informed by what may be called *comparative neuroprimatology*:

Q2: *How can the evolutionary quest be informed by studying brain, behavior, and social interaction in monkeys, apes, and humans?*

This does not deny the importance of neuroethology more generally (e.g., the study of brain mechanisms related to birdsong to provide models of vocal learning; see Fitch & Jarvis, 2013, for a review) or the use of genetically modified mice to explore the genetic underpinnings of (nonlinguistic) neural mechanisms. But the argument here is that to understand the specifics of the evolution of *human* brain and behavior we need to hypothesize properties of the brain and behavior of our last common ancestor with our closest living relative, the chimpanzee (LCA-c), or our more distant ancestor shared with the macaque (LCA-m), or other monkeys and apes. This in turn requires one to compare behavior and brain mechanisms across these primate species, sharpening one's analysis of each class of behaviors by analyzing similarities and differences between two or more species—all the better to understand what is and is not uniquely human about the use of language.

Q3: *How can computational modeling advance these studies?*

The final hypothesis is that these lessons can be enriched by development of a *computational comparative neuroprimatology*: using computational modeling to assess the contributions of specific brain regions or neural circuitry to a class of behaviors in one species, proceeding thence to offer a more detailed analysis of how *neural* similarities and differences across species can enrich our understanding of *behavioral* similarities and differences. We may see this last arena as an important subclass of *computational neuroethology*, providing a particular perspective on ethology, the study of animal behavior, more generally. For example, Arbib (2003) brings computational neural models of frog visuomotor coordination and rat navigation into the mix. Nonetheless, if our aim is to understand the evolution of the human brain, then comparison with the brains of monkeys and apes is paramount. The study of frogs, songbirds, mice, and rats enriches our understanding of potential mechanisms, but nonetheless will play a secondary role.

Mirror Neurons and Systems

As is well known, mirror neurons found in the macaque are active both when a monkey performs an action and observes

another (monkey or human) perform a similar action. Lacking the ability to make a systematic study of neurons in the human brain, people conducting human brain imaging have dubbed a human brain region a *mirror system* (or *mirror mechanism*) for a class of actions if it is significantly more active when the subject is either observing actions of that class or executing actions of that class as compared, in each case, to some control task. Some authors place great emphasis on the role of mirror neurons or mirror systems in a range of behaviors. As I have argued elsewhere (e.g., Arbib, 2012, pp.138–146) such emphasis may ignore the fact that mirror systems play their role only by virtue of their interactions with brain regions “beyond the mirror.” Unfortunately, it has become fashionable to “throw the baby out with the bathwater”—turning from a critique of excessive claims for mirror systems considered in isolation to excessive claims for the irrelevance of mirror systems to (social) cognitive neuroscience in general. It is a virtue of computational comparative neuroprimatology that it forces us to take a systems view of the role that mirror neurons do or do not play in some larger system. Given our concern with evolution, it is no surprise to find that modeling the differing roles of mirror neurons in different species (e.g., whether or not the species has imitation—and if so, of what form—or language) may reveal more differences “beyond the mirror” than in the mirror neurons themselves.

The Mirror System Hypothesis

The mirror system hypothesis (MSH) provides one approach to framing the evolution of the language-ready brain. MSH is based on comparisons of monkey, ape, and human praxis and communication across many studies spanning from Arbib and Rizzolatti (1997) and Rizzolatti and Arbib (1998) to *How the Brain Got Language* (Arbib, 2012). MSH got its name because Rizzolatti observed that area F5 of the macaque (the classic site for mirror neurons for manual actions) is homologous to (part of) Broca's area, classically thought of as an area for the production of language. My awareness of the finding of Poizner, Klima, and Bellugi (1987), that lesions of Broca's area in deaf signers induce a form of aphasia akin to the effect on spoken language of those with normal hearing, then led to the hypothesis that mirror neurons might be at the heart of language *parity* (that the hearer can often get the meaning of the speaker via a system that has a mirror mechanism for gestures at its core), and that manual gestures may have led vocal gestures in the evolution of the language-ready brain. Indeed, macaques and chimpanzees have manual dexterity and the ability to acquire new manual skills, but are incapable of vocal learning.

With this background, I can next outline the progression asserted by MSH, noting just a few of the debates related to its claims. Note that the terms *mirror neuron* and *mirror system*

do not occur in this outline. It is only when we seek to model the neural mechanisms (as briefly sampled in the next section) that our analysis assesses the interactions between mirror systems and neurally localized subsystems beyond the mirror. It was for this reason that I labeled a response to my critics (Arbib, 2013) as “Complex Imitation and the Language-Ready Brain” to emphasize that MSH has complex imitation as the hinge in the emergence of the language-ready brain, and thus MSH may be of value even to those who believe that mirror neurons are in some sense mythical (Hickok, 2014). Here, then, are the subhypotheses that constitute MSH. Each seems worthy of continued investigation, whatever the fate of the overall framework.

- LCA-m had skill learning for manual tasks, but little skill at imitation. Communication was restricted to an innate set of vocal calls.
- LCA-c had skill learning and *simple imitation* for manual tasks, an innate set of vocal calls, and also a communicative repertoire of manual gestures, at least some of which are novel results of social interaction, with ontogenetic ritualization being one mechanism for this. The adjective *simple* here merely suggests that it lacks ingredients that are a crucial part of human imitation, which is therefore called “complex.” The key difference is the “add on” in the human of the ability to attend to the “shape” of movements, as well as to form deeper hierarchies of motor skills (see Arbib, 2012, Chapter 7, for an extended discussion). The characterization of the diverse forms of imitation exhibited by different species remains a wide-open topic, and the search for underlying neural mechanisms even more so.
- Multiple innovations on the evolutionary path from LCA-c to *H. sapiens* involved both biological changes and niche construction:
 - Complex imitation combined complex action recognition—the ability not only to recognize behaviors as composites of familiar actions but also to recognize that some components were variants of familiar actions—with the ability to use such an analysis to guide imitation of the observed behavior.
 - Pantomime (as an opportunistic, open form of communication) built on complex imitation by using a series of intransitive gestures to mimic the actions of some behavior as a way of drawing the observer’s attention (based on context and complex action recognition) to some object, agent, action, or event associated with that behavior. Russon (2016) claims that great apes are capable of complex imitation and pantomime, whereas MSH claims these abilities are unique to the human lineage. Indeed, Russon offers an impressive set of examples from gorillas and, primarily, from orangutans—reminding us that comparative primatology indeed distinguishes the capacities of chimpanzees, bonobos, gorillas, and orangutans. But my prime concern here is that perhaps Russon uses these terms more generously than I do, and we thus need a concerted effort to more formally characterize the differences between the capacities revealed in Russon’s examples and those I attribute to humans.
- But pantomime is both energetically costly to perform and highly ambiguous to interpret. *Protosign* emerged within a community as a set of conventionalized pantomimes, with conventionalization simplifying the performance and restricting the scope. Thus, similar pantomimes could yield different protosigns (e.g., differentially simplifying a protosign for *flying bird* to get distinct protosigns for *flying* and *bird*).
- Together with this, a process of fractionation began, whereby complex pantomimes and protosigns could be broken arbitrarily in such a way that different parts might come to symbolize different aspects of the original meaning. To take an anachronistic example, pantomimes for *opening door* and *closing door* might only overlap in an initial twisting of the hand corresponding to twisting the door handle. This shared component might then become conventionalized as the sign for *door* with the complementary pieces interpreted as *open* and *close*. The process of fractionation is complemented by the formation of primitive constructions, in this case, (*door*; *x*), whose existence invites the community to develop signs for “things that can be done to doors.” This is the first step in the complexification of protolanguages (Wray, 2000). The debate over the nature of protolanguage is very much an open one (Arbib & Bickerton, 2010).
- However, in the early stages of protosign development, protohumans—like all nonhuman primates—lacked the ability for any but the most rudimentary voluntary vocal control and learning. Nonetheless, facial and vocal expressions were part of the innate communicative repertoire, and auditory-motor coordination was vital to survival in a world of predators and prey as well as conspecifics. MSH claims that it was the existence of an open-ended (but still rudimentary) system for using and combining manual protosigns that provided the social and cognitive opening for vocal control and learning to become adaptive, selecting for the new brain mechanisms. Brain mechanisms and social practices could then develop from there in an expanding spiral in which protolanguages could exploit both voice and gesture. This claim on the relative staging of hand and voice in language evolution remains controversial (Aboitiz, 2013; Bornkessel-Schlesewsky, Alday, & Schlesewsky, 2016; Coudé, 2016) and demands further research (Arbib, 2013).
- Finally, MSH claims that the earliest *H. sapiens* had all these brain mechanisms in place as well as a constructed

niche that supported rudimentary protolanguage, but did not have language. Languages, in the sense (specified above) of communication systems endowed with rich and open-ended lexicons and grammars supporting a compositional semantics, were then the emergent fruits of cultural evolution and required no change in the genome, save, perhaps, the result of Baldwinian evolution to support increased fluency in production and comprehension.

The data supporting these claims and related brain modeling as of 2012 were integrated into the exposition of *How the Brain Got Language*. This 2012 version of MSH had both its supporters and detractors, as is evident in a special issue of *Language and Cognition*, edited by David Kemmerer, which contained a summary of the book, 12 commentaries from diverse disciplines, and my response (Arbib, 2013), which was similar to “Towards a Computational Comparative Neuroprimatology: Framing the Language-Ready Brain” (Arbib, 2016a), which set forth a number of issues for building on the 2012 state of the art, the 18 commentaries, and my response (Arbib, 2016b). In each case my responses were not intended to blindly support my group’s work to date, but rather—as the previous examples demonstrate—to clarify those areas of agreement and disagreement that point the way forward to further productive research linking computational comparative neuroprimatology to language evolution. As our understanding reacts to new data from multiple fields and the development of new insights from modeling, aspects of MSH have shifted, as they should, but the essential orientation of the research program has held steady.

Modeling

Related work has included a true exercise in computational comparative neuroprimatology: using modeling of how the macaque brain subserves visuomanual coordination (Fagg & Arbib, 1998), the recognition of other’s actions (Bonaiuto, Rosta, & Arbib, 2007; Oztop & Arbib, 2002), and the role of self-recognition in skill learning (Bonaiuto & Arbib, 2010) to ground a model of how chimpanzee brains could support the acquisition of novel gestures (Arbib, Ganesh, & Gasser, 2014), namely, *ontogenetic ritualization* (OR).

Tomasello and Call (1997) proposed OR as a means whereby (some) ape gestures could emerge:

1. Individual A performs praxic behavior X and individual B consistently reacts by doing Y.
2. Subsequently, B anticipates A’s overall performance of X by starting to perform Y before A completes X.
3. Eventually, A anticipates B’s anticipation, producing a ritualized form X^R of X to elicit Y.

Liebal (2016) offers a brief assessment of how this relates to other forms of ontogeny of ape gestures. However, some authors (e.g., Hobaiter & Byrne, 2011) argue that ape gestures are either “species typical” or produced by only a single individual, and deny the reality of OR in wild populations while admitting that it does occur in captive apes. Thus, whatever the solution to the still open empirical controversy, detailed modeling of the phenomenon remains pertinent. Just to pick one crucial ingredient of the modeling: Our model of the mirror system offers a computational account of mechanisms whereby (in the previous scenario) B can recognize A’s goal (to get B to do Y) earlier, and earlier in the performance of X. At first, A simply terminates performance of X once B’s initiation of Y is recognized (this time a function of the anticipatory nature of A’s mirror system). The crucial difference we posit between macaque and chimpanzee is that the chimpanzee can come to recognize proprioceptive cues for this early termination whereas the monkey cannot: Thus, the chimpanzee can ritualize X to get X^R based on a proprioceptive rather than only a visual goal. The model of Arbib et al. (2014) for the emergence of gestures is relatively simple and is (loosely) associated with diffusion tensor imaging of macaques, chimpanzees, and humans (Hecht et al., 2012).

Unfortunately, new NIH guidelines will inhibit or terminate U.S. research on the brains of living apes, and there are almost no useful data on the functioning of the human brain at the level of detailed neural circuitry—as distinct from the gross data afforded by, for example, brain imaging and data from neurological disorders. Fortunately, the rich database linking macaque neurophysiology and behavior continues to grow. How, then, are we to proceed with computational comparative neuroprimatology? In many cases—such as the reach to grasp and the control of eye movements—we assume that the underlying circuitry as charted in the monkey is relevant to filling in the details obtained from brain imaging and other human studies. Synthetic brain imaging (SBI) offers algorithms for averaging the synaptic activity revealed by model simulations of neural circuitry to predict region-by-region activity, and thus treat the models against brain imaging studies. The related methodology is to develop neural network models of human brain mechanisms for which one believes the relevant circuitry is similar to that revealed by animal (e.g., monkey) single cell neurophysiology, and then process simulation results at the neural network level to infer predictions that can be tested against data from human fMRI or other noninvasive measures (Arbib, Fagg, & Grafton, 2002). Related strategies are also available for testing models against human evoked response potential (ERP) data (Barrès, Simons, & Arbib, 2013).

To extend the approach to mechanisms such as those subserving language, for which nonhuman neural circuitry does not suffice, we use hypotheses about the evolution of brain mechanisms to suggest how macaque circuitry is modified and expanded upon in the architecture of the human

brain, then use the resultant model to make predictions for human brain imaging or for lesions. In the aforementioned chimpanzee model (Arbib et al., 2014), we applied this methodology to the comparison of macaque and chimpanzee brains with respect to communicative hand gestures. This model has another important feature: It introduces *dyadic brain modeling*, which focuses on what happens in the brains of two interacting agents, a dyad, where the actions of one influence the actions of the other, with both brains changing in the process. Luc Steels (e.g., Beuls & Steels, 2013; Steels, 2011) has used a simulation of embodied agents in evolutionary games in a fashion relevant to studies of (cultural) language evolution. Our innovation here is to provide the agents with “brains” based on prior work in brain modeling.

A clear challenge for computational work is to link the language game approach with the neuroprimatology approach—namely, to assess whether and how the representations and mechanisms assumed for agents in language games, which support the study of cultural evolution, including grammaticalization (Heine, 2016), are indeed available in the language-ready brain (perhaps after some prior ontogeny in an appropriate cultural niche). If not, as I suspect is the case, I see the reconciliation of brain modeling with AI-defined software of the agents to be a major and exciting research challenge.

To close this section, consider an aspect of language that has gone unaddressed by MSH, namely, the crucial role of *turn taking* in conversation. Moulin-Frier, Sanchez-Fibla, and Verschure (2015) developed a computational model, in the interacting AI agents mode, where turn-taking behavior emerges in populations of agents needing to maintain group cohesion for a survival purpose (e.g., because they are not adapted to survive in isolation). As an example, they consider marmoset monkeys living in a dense forest preventing visual contact. A way to maintain group cohesion there is through vocalizations that convey information about the presence and state of each group member. However, if several agents are vocalizing at the same time, these vocalizations will interfere, making agent identification and thus group cohesion harder to realize. This provides the adaptive pressure favoring the emergence of a turn-taking strategy.

What is most relevant here is the mention of marmosets, which indeed exhibit cooperative vocal communication by taking turns, with experimental data, suggesting that the behavior is learnt during infancy (Takahashi et al., 2015). Moreover, Takahashi, Narayanan, and Ghazanfar (2012) developed a computational model based on the interactions among three neural structures (*drive*, *motor*, and *auditory*) with feedback connectivity inspired by published physiological and anatomical data. They fitted the model to the temporal dynamics of spontaneous vocalizations produced by isolated marmosets, and then tested the model for its ability to predict the structure of vocal exchanges between two marmosets. This

is a long way from the modeling of detailed interactions between parietal and premotor circuitry of the macaque described earlier, but it does suggest a path for extending those models toward dyadic interaction and the auditory system. However, vocal turn taking is not heard in macaques. How, then, can we integrate our macaque models with models of turn taking in such a way that we can use data from neuroanatomy and neurophysiology to assess how differences in the brains of macaques and marmosets can explain why marmosets exhibit turn taking, whereas macaques do not?

Characterizing the Language-Ready and Language-Using Brain

A theory of evolution needs a clear characterization of what it is that evolved. Because certain changes in the human brain, adaptive for reading, occur at the ontogenetic time scale within a niche constructed by cultural evolution (Dehaene et al., 2010, explore how learning to read changes the cortical networks for vision and language), Colag  (2016) supports the distinction between the language-ready and the language-using brain, suggesting that analogous cultural and ontogenetic processes formed new functional and anatomical neural processes necessary to support full-blown human language as distinct from the protolanguages we posit for early groups of *H. sapiens*. The task of our approach to the study of language evolution based on comparative neuroprimatology, then, is to better understand what aspects distinguish the human brain from ape or monkey brains as a basis for assessing what makes only the human brain language ready.

All this poses the challenge of titrating biological and cultural evolution. However, where we can use fMRI to compare (at a coarse level) the brains of humans who can and cannot read, we live in a society in which all adults with “normal” brains use language. Nonetheless, there is a large (and changing) population of humans who do not have language—namely, infants. Thus, a long-term strategy for characterizing the language-ready brain would combine studies of language acquisition and other forms of cognitive development with longitudinal brain imaging and ERPs (evoked response potentials across the scalp) to seek to better understand and model what it is about the infant brain that enables the child’s embodied interaction with others to support the acquisition of diverse skills, including those for reading and language (cf. Dehaene-Lambertz, this issue). Hints of how this might work come from computational models of language acquisition (see MacWhinney, 2010, for the introduction to a special issue on computational models of child language learning), though few address what capabilities of the brain are needed to support the posited learning processes. Models of how the child learns to grasp, recognize affordances, and recognize the actions of others (e.g., Bonaiuto & Arbib, 2015; Bonaiuto et al., 2007;

Oztop, Bradley, & Arbib, 2004, from my own group) are, if MSH has any validity, also relevant—and note, for example, the observations of Caselli, Rinaldi, Stefanini, and Volterra (2012) on how the early actions and gestures of the child relate to its word comprehension and production. All this motivates the need, complementing what has gone before, for new approaches to computational neurolinguistics—how the language-ready brain functions when it has become a language-using brain. A clearer view of how circuitry in each brain region contributes to the production and comprehension of language could help focus the search for the phylogenetic and then ontogenetic processes that shape this outcome.

We can start with the crude overview shown in Fig. 1. The “mirror systems” are posited to contain not only mirror neurons but also other neurons so that they can mediate both production and recognition of the constituent actions. The starting point for the overall structure of Fig. 1 is that the visual system has both a dorsal path (“up” through parietal cortex) and a ventral path (“down” through inferotemporal cortex) from primary visual cortex to frontal cortex, and that studies of brain-damaged human patients suggests that

- *the ventral visual stream is a “what” system* (recognizing what objects are in a scene and their *approximate* spatial relationship can serve as the basis for planning a course of action), while, given a plan of action
- *the dorsal visual stream is a “how” system* (passing detailed information about, e.g., the shape and disposition of objects involved to the motor systems that control the actions).

The top and bottom boxes together constitute a conceptual model (consistent with, but much less detailed than, the computational models mentioned earlier) of the brain systems

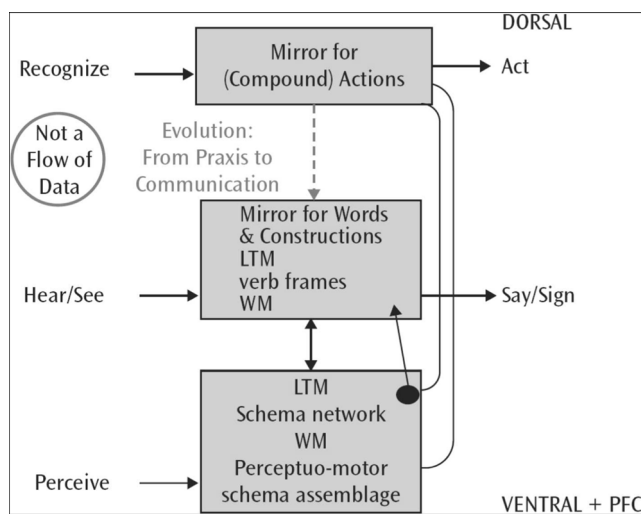


Fig. 1 Words as signifiers (articulatory actions or manually produced signs) link to signifieds (schemas for the corresponding concepts), not directly to the dorsal path for actions (Arbib, 2010).

common to monkeys, apes, and humans for visual control of manual actions. The ventral system maintains an assemblage of perceptual and motor schemas relevant to ongoing interactions with the physical and social environment; the dorsal system administers the detailed parameters necessary for successful execution of the motor schemas. Then, extending the basic scheme for single actions and words (more easily said than done!), we employ complex imitation as well as planning to lift execution and observation from single familiar actions to novel compounds, and similarly lift words to more complex utterances via the use of constructions (the details of this are beyond the scope of this brief review).

What may be surprising is that the arrow linking the “Mirror for Actions” to the “Mirror for Words” in Fig. 1 expresses an evolutionary relationship, not a flow of data. MSH sketches an account whereby the middle box of Fig. 1 evolves from the top box (incorporating its relation to the ventral pathway), tracing the evolution of protosign and thence protospeech via a process of conventionalization, which creates a class of *communicative* actions separate from *praxis* actions. (Note that the functional separation of the top two boxes is agnostic as to whether or not these functions are anatomically segregated.) This separation of *communicative* actions is reinforced when we note that words may serve diverse grammatical roles beyond that of verbs describing actions in the speaker’s repertoire. Rather than a direct linkage of the dorsal representation of the action to the dorsal representation of the articulatory form, we have two relationships between the dorsal pathway for the “Mirror for Actions” and the schema networks and assemblages of the ventral pathway and prefrontal cortex. The rightmost path of Fig. 1 corresponds to the connections whereby inferotemporal cortex and prefrontal cortex can affect the pattern of dorsal control of action. The path just to the left of this shows that the dorsal representation of actions can only be linked to verbs via ventral schemas. This general scheme might be supplemented by direct connections in those special cases when the word does indeed represent an action in the person’s own repertoire, but there are scant data on how generally such connections might occur.

Of course, an animal’s awareness of the world can depend as much on audition and touch and other senses as on vision, and so a general model of animal behavior must integrate multiple senses in linking perception to action. However, for our present purposes, the key issue is that much of human language use is auditory–vocal rather than visual–manual. As already noted, one of the open debates in the evolution of language is whether the path from monkey-like vocalizations in LCA-m via LCA-c to speech is direct (vocal control and learning evolved prior to the emergence of anything like a compositional semantics) or indirect (with protosign providing semantic scaffolding for the emergence of vocal mechanisms to serve meaningful speech). However, our concern in

this section is with more fully characterizing the modern human language-using brain, and here auditory–vocal language is dominant (though accompanied by cospeech gestures, and with a brain that can equally well learn to handle the signed languages of the deaf). It is thus of great interest that the primate brain also contains dorsal and ventral streams in the *auditory* system, as first revealed in work on macaque auditory neurophysiology (e.g., Rauschecker & Scott, 2009; Romanski et al., 1999). The most famous analysis of these pathways in neurolinguistics is that postulated by Hickok and Poeppel (2004, and see Hickok’s contribution to this issue) in their analysis of cortical stages of speech perception. The early stages involve auditory fields in the superior temporal gyrus bilaterally (although asymmetrically), but this cortical processing system then diverges into two streams:

- *A dorsal stream mapping sound onto articulatory-based representations*, which projects dorsoposteriorly and ultimately projects to frontal regions. This network provides a mechanism for the development and maintenance of “parity” between auditory and motor representations of speech. It thus corresponds to the auditory processing posited for recognition and production of words-as-actions in the middle box of Fig. 1.
- *A ventral stream mapping sound onto meaning*, which projects ventrolaterally toward inferior posterior temporal cortex (posterior middle temporal gyrus) that serves as an interface between sound-based representations of speech in the superior temporal gyrus (again bilaterally) and widely distributed conceptual representations. It thus augments the bottom box in Fig. 1 by tracing the auditory path whereby perception of a word-as-signifier may access perceptual schemas that may then affect the updating of perceptuomotor schema assemblages.

The scheme of Hickok and Poeppel, then, offers an essential ingredient in refining the conceptual model of Fig. 1 to attend to data on the role of auditory processing in speech perception, but does so only at the word level. A conceptual model of the role of the *auditory* dorsal and ventral streams in sentence comprehension (the B&S model; Bornkessel-Schlesewsky & Schlesewsky, 2013) has been developed as a cortical instantiation of the extended argument dependency model (eADM; Bornkessel & Schlesewsky, 2006). Their approach is described at some length in Section 4 of Arbib (2015a), which also describes attempts to offer a computational version of construction grammar that clarifies how perception of a visual scene may be linked to utterances describing it (Barrès & Lee, 2014). Meanwhile Dominey and his colleagues (starting with Dominey & Inui, 2009) have developed computational models of how corticostriatal interactions may be involved in language processing. An ongoing challenge in computational neurolinguistics is to assess how future

theorizing can develop an integrated perspective while addressing the commonalities and divergences of these and other (conceptual) models, such as those of Friederici (2011) and Hagoort (2013).

Like most workers in neurolinguistics, Hickok and Poeppel, and Bornkessel-Schlesewsky and Schlesewsky explore the roles of the auditory pathways in the comprehension of words and sentences, but neither addresses the mechanisms of articulation of words and they are silent on the use of hands and other effectors in sign language, and on the way in which vision enters into both our praxic and communicative interaction with the world, whether or not our language is spoken. Moreover, none of the efforts reviewed here assess the most important context in which language emerges, namely in conversation (Garrod & Pickering, 2004, 2009; Pickering & Garrod, 2013).

In summary, then, we have seen that computational comparative neuroprimatology, research on the evolution of the language-ready brain, and neurolinguistics are all in a state of flux, but I have outlined strategies (some more well-formed than others) for better linking and mutually calibrating these three fields. In particular, this section has served to open up an assessment of the capabilities of the language-using brain more generally than any focused on speech perception alone and thus sets new challenges for hypothesizing the biological and cultural processes underlying the evolution of the language-ready brain.

Acknowledgment This material is based in part on work supported by the National Science Foundation under Grant No. 0924674 and Grant No. BCS-1343544 “INSPIRE Track 1: Action, Vision and Language, and their Brain Mechanisms in Evolutionary Relationship” (Michael A. Arbib, Principal Investigator).

References

- Aboitiz, F. (2013). How did vocal behavior “take over” the gestural communication system? *Language and Cognition*, 5, 167–176.
- Arbib, M. A. (2003). Rana computatrix to human language: Towards a computational neuroethology of language evolution. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 361(1811), 2345–2379. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14599323
- Arbib, M. A. (2010). Mirror System Activity for Action and Language is Embedded in the Integration of Dorsal & Ventral Pathways. *Brain Lang*, 112(1), 12–24.
- Arbib, M. A. (2012). *How the brain got language: The mirror system hypothesis*. New York, NY: Oxford University Press.
- Arbib, M. A. (2013). Complex imitation and the language-ready brain. *Language and Cognition*, 5, 273–312.
- Arbib, M. A. (2016a). Towards a computational comparative neuroprimatology: Framing the language-ready brain. *Physics of Life Reviews*, 16, 1–54.
- Arbib, M. A. (2016b). Primates, computation, and the path to language. Reply to Comments on “Towards a computational comparative

- neuroprimatology: Framing the language-ready brain". *Phys Life Rev*, 16, 105–122.
- Arbib, M. A., & Bickerton, D. (Eds.). (2010). *The emergence of proto-language: Holophrasis vs compositionality*. Philadelphia, PA: John Benjamins.
- Arbib, M. A., Fagg, A. H., & Grafton, S. T. (2002). Synthetic PET imaging for grasping: From primate neurophysiology to human behavior. In F. T. Somer & A. Wichert (Eds.), *Exploratory analysis and data modeling in functional neuroimaging* (pp. 231–250). Cambridge, MA: MIT Press.
- Arbib, M. A., Ganesh, V., & Gasser, B. (2014). Dyadic brain modeling, ontogenetic ritualization of gesture in apes, and the contributions of primate mirror neuron systems. *Philosophical Transactions of the Royal Society B*, 369(1644), 20130414.
- Arbib, M. A., & Rizzolatti, G. (1997). Neural expectations: A possible evolutionary path from manual skills to language. *Communication and Cognition*, 29, 393–424.
- Barrès, V., & Lee, J. Y. (2014). Template construction grammar: From visual scene description to language comprehension and agrammatism. *Neuroinformatics*, 12(1), 181–208.
- Barrès, V., Simons, A., & Arbib, M. A. (2013). Synthetic event-related potentials: A computational bridge between neurolinguistic models and experiments. *Neural Networks*, 37, 66–92. doi:10.1016/j.neunet.2012.09.021
- Beuls, K., & Steels, L. (2013). Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PLoS ONE*, 8(3), e58960. doi:10.1371/journal.pone.0058960
- Bonaiuto, J. J., & Arbib, M. A. (2010). Extending the mirror neuron system model, II: What did I just do? A new role for mirror neurons. *Biological Cybernetics*, 102(4), 341–359. doi:10.1007/s00422-010-0371-0
- Bonaiuto, J. J., & Arbib, M. A. (2015). Learning to grasp and extract affordances: The integrated learning of grasps and affordances (ILGA) model. *Biological Cybernetics*, 109(6), 639–669. doi:10.1007/s00422-015-0666-2
- Bonaiuto, J. J., Rosta, E., & Arbib, M. A. (2007). Extending the mirror neuron system model, I : Audible actions and invisible grasps. *Biological Cybernetics*, 96, 9–38. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17028884
- Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4), 787–821. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17014303>.
- Bornkessel-Schlesewsky, I., Alday, P. M., & Schlesewsky, M. (2016). A modality-independent, neurobiological grounding for the combinatorial capacity of the language-ready brain: Comment on “Towards a computational comparative neuroprimatology: Framing the language-ready brain” by M. A. Arbib. *Physics of Life Reviews*, 16, 55–57.
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Reconciling time, space and function: A new dorsal–ventral stream model of sentence comprehension. *Brain and Language*, 125(1), 60–76. doi:10.1016/j.bandl.2013.01.010
- Caselli, M. C., Rinaldi, R., Stefanini, S., & Volterra, V. (2012). Early action and gesture “vocabulary” and its relation with word comprehension and production. *Child Development*, 83(2), 526–542.
- Colagè, I. (2016). The cultural evolution of language and brain: Comment on “Towards a computational comparative neuroprimatology: Framing the language-ready brain” by M. A. Arbib. *Physics of Life Reviews*, 16, 61–62.
- Coudé, G. (2016). On the role of emerging voluntary control of vocalization in language evolution: Comment on “Towards a computational comparative neuroprimatology: Framing the language-ready brain” by M. A. Arbib. *Physics of Life Reviews*.
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A.,...Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, 330(6009), 1359–1364. doi:10.1126/science.1194140
- Dominey, P. F., & Inui, T. (2009). Cortico-striatal function in sentence comprehension: Insights from neurophysiology and modeling. *Cortex*, 45, 8, 1012–1018
- Fagg, A. H., & Arbib, M. A. (1998). Modeling parietal-premotor interactions in primate control of grasping. *Neural Networks*, 11(7/8), 1277–1303. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12662750
- Fitch, W. T., & Jarvis, E. D. (2013). Birdsong and other animal models for human speech, song, and vocal learning. In M. A. Arbib (Ed.), *Strüngmann forum reports, Vol. 10. Language, music and the brain: A mysterious relationship* (pp. 499–539). Cambridge, MA: MIT Press.
- Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4), 1357–1392.
- Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Science*, 8(1), 8–11. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14697397
- Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2), 292–304. doi:10.1111/j.1756-8765.2009.01020.x
- Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, 4. doi:10.3389/fpsyg.2013.00416
- Hecht, E. E., Gutman, D. A., Preuss, T. M., Sanchez, M. M., Parr, L. A., & Rilling, J. K. (2012). Process versus product in social learning: Comparative diffusion tensor imaging of neural systems for action execution–observation matching in macaques, chimpanzees, and humans. *Cerebral Cortex*. doi:10.1093/cercor/bhs097
- Heine, B. (2016). On the role of context in language evolution: Comment on “Towards a computational comparative neuroprimatology: Framing the language-ready brain” by M. A. Arbib. *Physics of Life Reviews*, 16, 77–79.
- Hickok, G. (2014). *The myth of mirror neurons: The real neuroscience of communication and cognition*. New York & London: W.W. Norton
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1–2), 67–99.
- Hobaiter, C., & Byrne, R. W. (2011). The gestural repertoire of the wild chimpanzee. *Animal Cognition*, 14, 745–767. doi:10.1007/s10071-011-0409-2
- Laland, K. N., Odling-Smee, J., & Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences*, 23(01), 131–146. doi:10.1017/S0140525X00002417
- Liebal, K. (2016). The ontogeny of great ape gesture — not a simple story: Comment on “Towards a computational comparative neuroprimatology: Framing the language-ready brain” by M. A. Arbib. *Physics of Life Reviews*, 16, 85–87.
- MacWhinney, B. (2010). Computational models of child language learning: An introduction. *Journal of Child Language*, 37, 477–485.
- Moulin-Frier, C., Sanchez-Fibla, M., & Verschure, P. F. M. J. (2015). *Autonomous development of turn-taking behaviors in agent populations: A computational study*. Paper presented at the IEEE International Conference on Development and Learning, Providence, RI, USA.
- Oztop, E., & Arbib, M. A. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics*, 87(2), 116–140. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12181587

- Oztop, E., Bradley, N. S., & Arbib, M. A. (2004). Infant grasp learning: A computational model. *Experimental Brain Research*, 158(4), 480–503. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15221160
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347.
- Poizner, H., Klima, E. S., & Bellugi, U. (1987). What the hands reveal about the brain. Cambridge, MA: MIT Press.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/19471271>.
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5), 188–194. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9610880>.
- Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, 2(12), 1131–1136. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10570492
- Russon, A. (2016). Reconsidering great ape imitation and pantomime: Comment on “Towards a computational comparative neuroprimateology: Framing the language-ready brain” by M. A. Arbib. *Physics of Life Reviews*, 16, 91–92.
- Steels, L. (2011). Modeling the cultural evolution of language. *Physics of Life Reviews*, 8(4), 339–356.
- Takahashi, D. Y., Fenley, A. R., Teramoto, Y., Narayanan, D. Z., Borjon, J. I., Holmes, P., & Ghazanfar, A. A. (2015). The developmental dynamics of marmoset monkey vocal production. *Science*, 349(6249), 734–738.
- Takahashi, D. Y., Narayanan, D., & Ghazanfar, A. A. (2012, November 7–9). A computational model for vocal exchange dynamics and their development in marmoset monkeys. Paper presented at the 2012 I.E. International Conference on Development and Learning and Epigenetic Robotics (ICDL).
- Tomasello, M., & Call, J. (1997). *Primate cognition*. New York, NY: Oxford University Press.
- Wray, A. (2000). Holistic utterances in protolanguage: The link from primates to humans. In C. Knight, M. Studdert-Kennedy, & J. Hurford (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 285–302). Cambridge, UK: Cambridge University Press.