DE GRUYTER MOUTON
DOI 10.1515/langcog-2013-0020 — Language and Cognition 2013; 5(2–3): 273–312

Michael A. Arbib

Complex imitation and the language-ready brain

Abstract: The present article responds to commentaries from experts in anthropology, apraxia, archeology, linguistics, neuroanatomy, neuroimaging, neurophysiology, neuropsychology, primatology, sign language emergence and sign language neurolinguistics on the book *How the brain got language: The mirror system hypothesis* (Arbib 2012). The role of complex imitation is discussed, and the distinction between protolanguage and language is emphasized. Issues debated include the role of protosign in scaffolding protospeech, the interplay between biological evolution of the brain and cultural evolution of the social interactions within groups, the relations brain mechanisms for action and language, and the question of when language first emerged.

Keywords: language evolution, mirror neurons, mirror systems, brain, primate communication, imitation, schema theory, pantomime, protosign, protolanguage, protospeech, sign language, aphasia, apraxia, human origins

Michael A. Arbib: Computer Science, Neuroscience, and the USC Brain Project, University of Southern California, Los Angeles, CA 90089-2520, USA. E-mail: arbib@usc.edu

1 Introduction

The commentaries from anthropology, apraxia, archeology, linguistics, neuroanatomy, neuroimaging, neurophysiology, neuropsychology, primatology, sign language emergence and sign language neurolinguistics make responding a daunting (and necessarily incomplete) task, but it has been exciting to see how these come together in defining strategies for strengthening and building upon the Mirror System Hypothesis (MSH).

Here I need to make the usual comment that neither macaques nor chimpanzees are ancestral to humans. Macaques and humans diverged from the last common ancestor of humans with macaques (LAC-m) some 25 million years ago but I shall (along with all the commentators) ascribe some but not all properties of the macaque brain to LAC-m without repeating the necessary caveats. Similarly for LAC-c and chimpanzees, who have evolved apart from us for 5 to 7 million years.

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

By contrast, it is accepted that knuckle walking evolved along the chimpanzee line since LAC-c, and was not part of the evolution of human bipedalism.

As described in the précis which introduces this issue of *Language and Cognition*, MSH structures the evolution of the language-ready brain in six stages:

Pre-Hominid: 1. A mirror system for grasping (LAC-m); 2. A simple imitation system for grasping (LAC-c)
Hominid Evolution: 3. A complex imitation system; 4. Pantomime; 5. Protosign: 6. Protospeech and multi-modal Protolanguage
and then adds
Cultural Evolution in Homo Sapiens: 7. Language.

In what follows, I will use the notation H123 for page 123 of *How the Brain Got Language*, and set a commentator's name in **bold** when referring to their commentary.

2 In defense of mirror neurons

The mirror system occurs explicitly only in Stage 1 above. The larger claim is that mirror neurons for grasping interact with mechanisms "beyond the mirror" to scaffold simple and then complex imitation in the manual domain – a social extension of dexterity. The arguments for subsequent stages build on complex imitation once it is in place, whether or not mirror neurons are necessary. For this reason, it might be better to call my (still evolving) theory the *Complex Imitation Hypothesis of the Language-Ready Brain* and one must then reiterate its important parallels with Merlin Donald's account of mimesis (H175, H248). But this does not mean that I have abandoned mirror neurons.

Zuberbühler decries claims that mirror neurons "are *the* [my italics] neural mechanism underlying a number of complex cognitive processes, many of which define human uniqueness," failing to note that I offer a far more nuanced view in which mirror neurons are part of a larger system (H138–46). He objects that "mirror neurons [may] respond to similar, related, or even different actions in the observation and execution condition," but broad versus limited congruence was always part of the story – one I relate to population coding (H123–4).

"If mirror neurons are not the product of natural selection, but an acquired and general feature of brain circuitry, then it is difficult to see why they should be given a privileged position in theories of language evolution." Why should being an "acquired and general feature of brain circuitry" be less relevant to language than the innate species-specific vocalizations, resistant to learning, that **Zuber-**

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

bühler studies? The MNS models (H128–135) show computationally that mirror neurons are ontogenetically acquired through adaptive circuitry that evolved to make such ontogeny possible. The ACQ model (H139–40) posits a general utility of mirror neurons for learning from one's own actions. It uses cats as an example, but does not argue that protocats evolved into language-bearing felines. I am comfortable with the view that "mirror neurons [may] turn out to be a general feature of vertebrate brains." The issue is whether the existence of mirror neurons *for manual action* did or did not support the path via complex imitation to pantomime and beyond.

In conclusion, **Zuberbühler** says he "cannot get so excited over the fact that some cells in the macaque brain show interesting firing patterns" – a hostile putdown of neuroscience, I would say.

Emmorey says that I propose "that a mirror system for grasping evolved into a mirror system that is not tied to praxic actions, but rather relates to actions of speaking (or signing)." Not quite. I claim that a mirror system for speaking (or signing) evolved "*atop*" a mirror system for grasping but I do hypothesize a mirror system for words-as-articulatory-objects, asserting that these are linked to a semantic system that is not part of the mirror system (Fig1.left). I would then insist that language and the semantic system co-evolve – there is no prior conceptual system *of modern richness* waiting to be expressed in language. A reminder: evidence for mirror systems (note plural) in humans is primarily from brain imaging, and their activation may reflect mirror neuron activity in some protocols but owe more to non-mirror neurons in others (H136).

Emmorey cites data suggesting that a mirror system for the actions of speaking or signing is not critical for either spoken or signed language processing, noting Hickok's arguments "contra mirror neurons." I find myself in partial agreement with them, but there's a lot to unpack here:

(i) We know that Broca's aphasia (which is not limited to damage to Broca's area) is associated more with agrammatism than with deficits in word comprehension. In response to **Dominey**, then, one has to ask how constructions (as distinct from the lexicon) are realized neurally. This is still being explored, but probably involves cooperation between prefrontal cortex and basal ganglia.

(ii) The dorsal-ventral division shown in Fig.1.left is reminiscent of that postulated by Hickok and Poeppel (2004) in their analysis of cortical stages of speech perception: a dorsal stream mapping sound onto articulatory-based representations, and a ventral stream mapping sound onto meaning. I argue that the ventral path alone can usually get from sound to meaning *but the dorsal path is critical if we are to acquire a novel word or imitate a novel accent*. Moreover, if there are problems of "load" (e.g. effects of noise), normals may mouth words to assist

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

Fig. 1: (Left) Words as signifiers (articulatory actions) link to signifieds (schemas for the corresponding concepts), not directly to the dorsal path for actions. The basic scheme is enriched by compound actions and constructions. LTM, long-term memory; PFC, prefrontal cortex; WM, working memory (H282). (Right) A variant of the Rothi et al. dual-route model of praxis. The key novelty of this model is that actions specified by both the direct and indirect routes (the latter via the praxicons) may be combined in the action buffer (H206).

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

understanding, whereas the word comprehension of those with Broca's damage drops off.

(iii) If one concedes that there is an MNS for human action, then one needs similar studies of when it is or is not necessary. Some studies of apraxia are discussed further below in connection with **Pazzaglia** and **Stasenko, Garcea, and Mahon**, but here the focus is on imitating facial expressions of emotion. Dapretto et al. (2006) presented children with faces expressing different emotions. Subjects either imitated or simply observed the faces. During observation, fMRI activity in the putative mirror system for facial action (stronger on the right) was seen in the typically developing group but this activity decreased with a measure of severity on an autism scale for an autism spectrum disorder (ASD) group However, there were no group differences in how well the children imitated facial expressions. A Disturbing Thought: A mirror system isn't important for recognition of facial expressions! However, children with ASD showed greater activity than did the typically developing children in regions modulated by visual and motor attention. Would a partial view of a face expressing emotion yield the full facial expression in normals but just the shown movement in autistics? Dapretto (personal communication) says the experiment has not been done, but she has seen data where an ASD child viewing the full face would develop the facial expression piecemeal rather than as a whole. The hypothesis still awaiting test, then, is that this mirror system provides a path for *empathizing with emotions* whereas the "autistic alternative" uses another path to "assemble motor features."

(iv) This connects with Fig.1.right (H201–208), with the indirect path implicating the mirror system in imitating or pantomiming known actions (linked to semantics) while the direct path assembles known features. My "theory of tweaks" suggests how the two paths may work together to learn novel actions. However, much work remains to be done to address a range of data on apraxia and specify subsystems which can be localized in the human brain.

(v) Where Buccino et al. (2004) take an either/or view for recognition of nonconspecific versus conspecific actions, I argue (H142) that the machinery for recognition of nonconspecific actions is also available as a dual pathway for recognition of conspecific actions. I argue that MNS would light up if we were asked to *imitate* a dog barking even though it does not for observation.

With all this, my current position is (non-controversially) that a motor system is needed to articulate words and that (apparently, controversially) a mirror system is needed to follow novel articulatory structure for imitation or in combating noise to better access semantic representations and, a *fortiori*, to acquire the ability to speak (or sign) the words of a language. Challenging questions remain for future research.

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

3 Learning from primate communication

H.Chap.3 suggests that gesture in apes may exhibit social learning absent in vocalization, supporting a view of the prime importance of manual gesture in language evolution. **Zuberbühler** usefully notes that chimpanzee vocal behavior is strongly mediated by social variables, and that individuals can use vocalizations to persuade and inform others, but offers no evidence for learning of novel vocalizations. He says my scenario is that language evolved in two stages. Manual-based protolanguage came first, followed by a vocal-based protolanguage that came second. But it's more subtle than that. I posit a spectrum of protolanguages with protosign getting a head start to open up semantics, but then posit an expanding spiral as each supports the evolution of the other – until, *historically though not biologically*, speech has become dominant. He suggests that speech evolved directly from vocal communication, with gestures playing a subsidiary role. It's possible, but for me the transition *pantomime → open semantics* is the tie breaker in favor of a crucial gestural scaffolding for the evolution of speech.

Zuberbühler says "Pantomiming is conspicuously absent [in apes], apart from isolated anecdotes." I agree, seeing this as post-LCA-c, unconvinced by Russon and Andrews (2010). He adds that "part of [Arbib's] strategy is to discount primate vocal behavior as irrelevant for questions of language evolution." Discount, yes; irrelevant, no. The puzzle is to link the medial call system to the lateral language system (H118–21) and I discuss the puzzle but do not claim to solve it.

He then reviews evidence for persuasion, inhibition and provision of information in great ape vocal communication. However, I do not see evidence for "assessment of the recipient's knowledge" as distinct from audience effects (H74) and effects of dominance relations, as in the work on baboons by Cheney and Seyfarth (2007). Evidence that chimpanzee vocal behavior is under considerable social control is not evidence that it is under "intentional control." The one result that seems truly challenging is that "wild Thomas langur males were observed when alarm calling to a tiger model. . . . the males continued to alarm call until every group member had responded with at least one alarm call, as if trying to ensure that all group members were aware of the danger," though this is based on observing behavior rather than inferring intentions and one would like to know more about the processes for checking "everyone has called."

Zuberbühler cites the flexibility that arises when some primates produce sequences of calls which can sometimes lead to changes in meaning but does not address my argument (H75–76) that this is not a precursor of syntax. As he observes, "it is not possible for a chimpanzee to explain to another what it had for

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

breakfast, but it can indicate the location of food, inform others about danger, or choose to remain silent if social conditions are unfavorable," but he remains silent on why this would be relevant to the emergence of languages with an open lexicon and rich set of constructions. My claim is that ape-like gesture opens the path to pantomime, which calls do not, but that a further step – the transition to protosign – was required to get us to "full" symbolization.

4 Learning from sign language

The commentaries by **Sandler** and **Emmorey** are centered on sign language, but first I need to address the misconception of **Aboitiz** that "[because] the articulatory organization of speech patterns is much faster than that of manual and body gestures, the vocalizer [can] generate highly complex messages in a brief time, a capacity that is much less obvious in hand signing." However, parallelism in sign allows it to support the same propositional rate as speech, as suggested by simultaneous translation of speech into sign (Padden 2000/01). Bellugi and Fischer (1972) found that a sign in American Sign Language (ASL) takes longer to produce than a spoken word, but that a proposition takes about the same amount of time to produce in either language. So-One Hwang (2011) found production rates in sentences taken from natural conversations of English, Korean, and ASL were consistent with the hypothesis of modality-independent time pressures for language processing.

Carol Padden is a co-investigator with **Sandler** of the Al Sayyid Bedouin Sign Language, ABSL, of H-Chap.12. She notes (personal communication, 2012): "The body has more articulators that can move at one time. [Compare **Sandler**, Figure 2.] . . . as [ABSL] matures, the parts of the body [used by signers] become synchronized and the speed of signing increases. . . . It's hard to appreciate how fast signers can get if you're not already a signer." Such observations suggest that the same central semantic system (the schema system of Figure1.left) is tapped at different *movement* speeds (but the same "*semantic rate*") by different articulatory systems.

Contra **Sandler**'s assertion, I do not propose continuity between holistic pantomime and signs: I propose (H-Chap.8) a discontinuity between "artless" pantomime (using variants of the available action repertoire to communicate) and protosign (conventionalizing pantomime and adding arbitrary gestures as a social symbolic system); and then suggest (H-Chap.10) how initially holophrastic protowords (signed or spoken) may have become fractionated in a way which simultaneously generated constructions which could put the pieces together again in novel combinations.

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

Sandler asserts that the modern ability to create new sign languages reveals the richness and plasticity of human cognition, and not an evolutionary stepping stone to speech, but I hold that the "plasticity of human cognition" required a language-ready brain and the "richness of cognition" requires social development in a language-rich culture. **Sandler** stresses the gulf between the grammar of spoken and sign languages. But consider the gulf even between spoken languages – history allows cultures to develop diverse means of communicating by speech as well as by sign. For example, English "Some crazy boys decided to make some wine to sell" translates in Navajo to

Ashiiké	**t'óó diigis**	**léi'**	**tółikaní**	**ła'**	**ádiilníil**	**dóó**	**nihaa**
boys	*foolish*	*certain*	*wine*	*some*	*we'll make*	*and*	*from us*

nahidoonih	**níigo**	**yee**	**hodeez'ą**	**jiní.**
it will be bought	*they saying*	*with it*	*they planned*	*it is said*

(Source: http://en.wikipedia.org/wiki/Navajo_language, downloaded January 3, 2013).

Compare **Sandler's** example: The Israeli Sign Language version of "The little dog that I found last week (over there) ran away" can be glossed as

DOG SMALL INDEX I FIND LAST WEEK INDEX **('there')** // ESCAPE

The double slash marks here separate the topic of the sentence from the comment. The indices are pointing signs used as referential loci in the system. Sign is especially rich in its use of peripersonal space, e.g. using a return to the location where a sign was made to name someone as a "pronoun" for that person specific to the current utterance. It is thus suggestive for MSH that Almor et al. (2007), using fMRI, found that reading pairs of sentences (visual coding of speech!) with repeated names elicited more activation than pronouns in the middle and inferior temporal gyri and intraparietal sulcus. They suggest that the latter activation is related to spatial attention and perceptual integration – which may relate to the use of signing space. Moreover, **Diessel** explores how pointing/deixis enriches the development of spoken language.

Sign language must extract "words" from visual input and output words as motor commands to face and hands; spoken language must extract words from auditory input and output words as motor commands to the vocal apparatus (with cospeech gestures possibly involving hands as well). Human brains have

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

widely distributed mechanisms which can build up neural codes for an immense lexicon that is connected to a rich store of neurally coded concepts widely distributed in the brain, as well as neural processes implementing a rich grammar which allows novel meanings to be extracted from sensory input and expressed in motor output. *But these widely distributed mechanisms have no direct knowledge of hands or faces or muscles – they simply process neural codes for input and output to allow us to carry on meaningful conversations and learn as we do so.* When seen from this point of view, the difference between English and a sign language is no greater than the difference between English and Navajo. Questions for study of evolution of the language-ready brain include

a) how do we evolve the sensory and motor interfaces for speech and sign, and
b) how do we evolve the central mechanisms for lexicon and grammar?

My claim is that the sensory and motor interfaces for face and hand (but not voice) came first, then the rudiments of protosign established the first iteration of the "protolanguage-ready brain," then the vocal apparatus evolved to exploit the central mechanisms, and thereafter both manual-based and voice-based contributions to protolanguage did double duty: (i) they contributed to refinement of their own sensory-motor interfaces to the (initially lexical then increasingly grammatic) central mechanisms; and (ii) each contributed to refinement of the central mechanisms in such a way that both benefited equally.

Sandler shows (as does my book) that sign languages have duality of patterning (H141) and that any connection between pantomime and sign may be indirect (H-Fig.2–4). Similarly, we often learn spoken words unaware of their etymology – but that does not deny that they have an etymology, and signs may (but may not) have an "iconic etymology" without that figuring in their use. She adds that sign languages have many commonalities with spoken language, including embedded clauses, morphological complexity and wh-movement.

Sandler sees it as a "kink" in MSH that the speech system seems so specialized as to have required an evolutionary process specific to speech, whereas the human hand is not specialized for sign language. But this is what MSH seeks to explain – protosign exploits prior circuitry for manual skills to provide scaffolding for evolution of articulatory control to serve flexible spoken communication.

Sandler notes my inclusion in pantomime of such actions as tracing the shape of an object with the hands. She sees this as entailing abstract symbolization and not reenactment, and further sees this as opposing my claim that pantomime opens up the rich semantics needed for language. I don't see the force of this argument. I see artless or "naïve" pantomime as a halfway house. MSH posits two major transitions: from imitation of praxic actions to (naïve) pantomime and

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

from pantomime to protosign. Indeed, one ABSL family uses a stereotyped version of peeling a banana as their ABSL sign for banana and another adds biting the top off the peeled banana. Or consider the AIRPLANE versus FLY issue (H220) of finding an arbitrary gestural tweak to distinguish concepts that might be denoted by the same pantomime. (Protohumans did not discuss airplanes, but consider the challenge of disentangling "bird" from "flying bird" that would allow one to meaningfully concatenate protosigns for "dead" and "bird." Perhaps missigning "dead" with "flying bird" was the proto-joke.)

"But as soon as the hands represent something other than the hands, such as the feet for 'jump', the eyes for 'see' . . . we are talking about symbolization, not pantomime." Indeed, the extension of imitation from using the same effector to using a corresponding motion of another effector is noted as an MSH substage. Perhaps what **Sandler** sees as a minus is actually a plus – as pantomime becomes extended it can be thought of as more symbolic. I don't see a strict dichotomy here. Note how easily children play with, e.g. moving the fingers for walking, or using the hand to emulate an airplane. For MSH, the key distinction is between a performance understandable by a naïve observer, and a performance that relies on the familiarity of the observer with the conventionalization involved. Lesions which yield sign language aphasia may preserve pantomime (H305), and we may link brain regions so damaged to the long-ago evolution of mechanisms which first supported protosign.

And consider **Sandler's** "the hands represent[ing] . . . meandering mental activity for 'dream' . . . Huge proportions of any sign language lexicon are grounded in this type of symbolization, in which the hands do not represent the hands, and iconicity is exploited symbolically and *metaphorically*." Her Figure 4 indicates a hand tracing a meandering path next to the head, but this entails a theory of mind which links the mind to the head rather than the heart. This is not something to be pantomimed by protohumans; it is the sign of someone acting within a language-rich environment with an advanced folk psychology. As I note [H320], the ABSL community always integrated deaf and hearing people in the same extended family; thus, the speakers who already knew how to express something in Arabic would be motivated to try to convey these same ideas by (not so naïve) pantomime supporting the development of increasingly conventionalized gestures. However, I do think metaphor may have been operative even at the protosign stage. One has an idea one wishes to express. One has no pantomime or protosign for it. One thinks of a concept which shares some features but not others and uses that instead. The context allows the observer to get the message. This may be a "once off" or may establish a new frozen metaphor that extends the expressive lexicon. As grammar emerges as we move along the protolanguage spectrum to languages, so may metaphors rise to the phrasal level. And on to

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

idioms. In summary, **Sandler** offers us interesting insights into modern sign language with a modern (though hearing-deprived) brain, but it is a mistake to use the richness of modern sign language against the plausibility of conscious use of pantomime providing the underpinning for early protosign on the path to both spoken and signed languages.

Sandler stresses that ABSL began without phonology, citing work I emphasize (H313–314) – and we agree that phonology may well have been an emergent as protolanguages developed (H269–271). However, she says it did not begin with pantomimic holophrases: "Its first linguistic building block is the word." This is a key issue, offered as a critique of my hypothesis (H-Chap.10) that the earliest protowords were holophrastic. I have two responses: (a) The ABSL community contains Arabic speakers who already use words and so have a habit they will try to share with their deaf relatives. (b) Some ABSL "words" are more akin to holophrastic protowords – recall *peeling a banana* for BANANA – than to "modern" words, with refinement occurring across the generations.

The only data **Sandler** has of a first generation ABSL signer is a single videotape of one telling a story. "[M]ost of his utterances consist of one or two words . . . he uses a sign for HIT which is not derived from the pantomimic form in any obvious way." But if an utterance consists of one word, isn't it a holophrase? The fact that the sign is later refined for a verb related to the original holophrase does not change this. And if HIT was not derived from pantomime, how was it derived? For example, we know that some signs of Nicaraguan Sign Language were imported from American Sign Language (H306) whereas others were fractionated from pantomimes (H303). In any case, **Sandler's** discussion of successive strata adding articulators is worth noting (her Table 1), and recalls both Padden's comment on skilled signing, and the developmental trajectory of infants learning to speak (H194) – it seems that skillful coordination of articulators is the product of becoming a member of a skilled community, or an emergent of the development of such a community.

In conclusion, **Sandler** overplays the difference in articulators for speech and sign, ignoring the issue of the evolution of the modality-independent central mechanisms that make this deployment possible – the language-ready brain – and underplays the sophistication of the cultural milieux in which new sign languages emerge.

Turning to the neurolinguistics of sign, **Emmorey** notes that Knapp and Corina (2010) conducted a conjunction analysis with six deaf ASL signers and found some overlapping activation in Brodmann Area (BA) 44 – part of Broca's area – for sign perception and production, but found other areas active as well. Indeed, they explicitly address MSH and suggest that if a unitary mirror system mediates the observation and production of both language and non-linguistic

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

action, three predictions can be made: (1) damage to the human mirror neuron system should non-selectively disrupt both sign language and non-linguistic action processing; (2) within the domain of sign language, a given mirror neuron locus should mediate both perception and production; and (3) the action-based tuning curves of individual mirror neurons should support the highly circumscribed set of motions that form the "vocabulary of action" for signed languages. They evaluate data from the sign language and mirror neuron literatures and find that these predictions are only partially upheld. But (1) does not follow from MSH (my above comments show that given loci may be differentially activated as a function of task) and (3) follows from an erroneous "one neuron, one action" view of mirror neurons that is prevalent in the literature but which I explicitly qualify in emphasizing population codes (H124). But there are still no adequate computational models of brain mechanisms of language processing. These need to be developed so we can see how best to address the data Knapp and Corina review.

Emmorey et al. (2010) used fMRI to study deaf signers and hearing individuals passively viewing video clips of pantomimes as well as ASL verbs that were rated as meaningless by non-signers. For hearing non-signers, both pantomimes and ASL verbs (which were meaningless for them) strongly activated the "mirror system" (i.e. fronto-parietal cortex) for human action but for deaf signers there was no activation in the mirror system during the perception of pantomimes whereas activation was found only in Broca's area during the perception of ASL verbs. **Emmorey** concludes that "the lack of activation within the mirror system for deaf signers [does] not support an account of human communication that depends upon automatic sensorimotor resonance between perception and action." But I do not use the word "resonance" in this way, and elsewhere have rejected this use of the term; and note the alternate paths of Figure 1.Left.

I agree with **Emmorey** in distinguishing conventionalized manual gesture from speech, and in noting the importance of McNeill's study of co-speech gestures (H39) – and her intriguing mention of Sandler's study of co-sign gestures with the mouth – and the challenge of understanding why or how this highly synergistic relationship between gesture and speech emerged from the expanding spiral between protosign and protospeech. (I could say more on sign and cosign, but that would be going off on a tangent.) An ad hoc reply is that a distributed, shared mechanism for semantic and other content can be tapped in diverse ways to control the expression of semantic and other content through voice, face and/or hands. Depending on one's culture, one or more of these outputs may follow (more or less) the well-formedness constraints of a language and the multiply interconnected pathways in the brain can mold these shared mechanisms in terms of the culture's form and content of expression. However, the very plasticity

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

of the human brain that supports cultural evolution ensures that the readiness of the brain to support the linkage between semantics and communication is neutral as to whether a human comes to employ spoken language, sign language or both. **Emmorey** states that MSH "would predict that the remnants of the gestural origins of language (i.e. pantomimes and modern protosigns) should co-occur with speech, but they do not." However, I would see co-speech gestures as the "remnants" (perhaps better: "they contains traces of remnants") that most often co-occur with speech without denying that pantomimes and modern protosigns form a complementary "remnant." Is this any more surprising than saying that a person who can speak both French and English rarely (but not impossibly) combines both in a single sentence without denying that both contain (rather than "are") remnants of Latin?

Emmorey further asserts that MSH is inconsistent with the properties of co-speech and co-sign gesture: "gesticulations of speakers are highly integrated with speech, but pantomimes and modern protosigns (conventional gestures) . . . do not co-occur with speech. Further, . . . [s]igners produce global, imagistic gesticulations with their mouths and bodies simultaneously while signing with their hands. The expanding spiral of protosign and protospeech does not predict the integrated and co-expressive nature of modern gestures produced by signers and speakers." I don't see the force of this. MSH roots language in a multi-modal system. It is a historical fact that most modern humans rely most on the voice for conventionalized communication whereas many deaf people place the major load on the hands, but nothing in MSH suggests that either convention should inhibit the expressive use of other effectors. Further, it is simply false that "conventional gestures . . . do not co-occur with speech." Normally, they don't need to, but someone who says "I called her on the phone and . . ." may well emit the conventional gesture for *phone* while speaking; and someone saying "I caught a fish that was yay big" will spread the hands to indicate the size. Moreover, the role of unconventionalized pantomimes in MSH is as a predecessor for protosign which is a predecessor for language. If indeed "naïve" pantomime or protosign rarely (not never) co-occurs with language use, this no more argues against MSH than saying that "humans do not crawl while they are walking" contradicts any notion that we evolved from quadrupeds.

A further speculation: Just as language integrates "ancient" medial systems and "modern" lateral systems to integrate affective and propositional content (H121), so can we consider pantomime as being at the root of two brain systems: a more ancient one that stays close to unconventionalized pantomime and a more modern one that evolved via protosign and protospeech. The former can then be expressed by motor systems that are not currently dominated by the language-expressive effectors. What is intriguing about Sandler's observation of co-sign is

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

that the implicit notion of "oral pantomime" gives another route for the oro-facial system to migrate toward protolanguage even before articulate vocalization emerged.

5 Vocalization and gesture in the evolution of the language-ready brain

Aboitiz asks "how did auditory-vocal circuits come to dominate a robust gesture-dedicated circuit, out of nearly nothing?" But he conflates the biological evolution of a brain for which hand and voice are equally flexible for communication, and cultural evolution in which speech becomes the chosen medium for almost all hearing societies, though gesture remains a major complement to speech. The right question for biological evolution might rather be "How did auditory-vocal circuits come to be on a par with a robust visuo-manual circuit when it appears that the latter was far better developed in both LCA-m and LCA-c?"

Aboitiz and **Fogassi, Coudé and Ferrari** (henceforth **FC&F**) emphasize pathways in the monkey that may be homologous to pathways for vocal control in the human, rejecting my doctrine of the expanding spiral of protosign and protospeech. But neither addresses my key argument, namely that an early form of protosign had to evolve atop pantomime as a manual-based communication system to break through the fixed repertoire of primate vocalizations to yield the *open-ended semantics* that rendered complexification of vocal control adaptive. I stress five parts of my account that are relevant here:

i) Monkeys do not have the flexible control of patterns of vocalization and articulation needed for speech (H74–75). **Zuberbühler** shows that chimpanzee vocalization is akin to that of monkeys. We may thus postulate that the repertoire of manual actions for both LCA-m and LCA-c is vastly greater than their expressive repertoire of vocal actions.

ii) Even for humans, "naïve" pantomime (characterized at H237) is more expressive than onomatopoeia. The same would hold even more strongly for descendants of LCA-c according to (i). Thus, a brain that could support pantomime as a conscious strategy could access an open semantics of a kind denied to humans and chimpanzees.

iii) However, pantomime is slow and ambiguous, favoring brains that could conventionalize and "annotate" pantomimes to yield an early-stage system of protosigns, as manual gestures.

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

iv) Once protosign (as distinct from pantomime) gets started, it establishes the cognitive machinery that can then begin to develop protospeech, perhaps initially through the utility of creating novel sounds to match degrees of freedom of manual gestures (rising pitch could represent an upward movement of the hand). This provides the adaptive pressure for increased control over the vocal apparatus. This control could then co-evolve with an ability for vocal imitation in attracting prey and confusing predators and with onomatopoeia to yield other vocal gestures not linked to manual gestures. Over time, an increasing number of symbols would have become vocalized, freeing the hands to engage in both praxis and communication as desired by the "speaker" (H235).

v) And thereafter, protosign and protospeech could co-evolve in an expanding spiral which (to make explicit a point perhaps underemphasized in the book) would involve differing mixes of gene-culture coevolution.

FC&F propose that vocalization in non-human primates (presumably descended from LCA-c, and now extinct) could have reached a partial voluntary control that, in conjunction with gestures, could have had an active role in the emergence of the first voluntary forms of utterances, which they assert to be protospeech. My claim is that the data on macaque vocal control reviewed by **Aboitiz** and **FC&F** do not provide evidence against (i)–(iii) and actually strengthen the claim for (iv). I exclude a sign language prior to spoken language, and require only limited protosign to ground the expanding spiral. I agree fully with **Aboitiz** that "Modern speech requires a complex combination of peripheral modifications." but stress that no nonhuman primate has vocal imitation whereas apes do exhibit imitation of manual skills, and so argue that protosign-scaffolded control of the vocal articulators in fact provided the breakthrough necessary to support vocal learning.

Aboitiz argues

a) Complex vocal learning can be achieved without need of a voluntary hand grasping circuit.
b) There are rudimentary circuits in the monkey that can convey auditory information into Broca's region.

For (a), "can" is proved by birdsong, but our concern is with human evolution. Moreover, elaboration of birdsong serves male competition for mates and defense of territory without a compositional semantics. The issue is to assess what pressures drove the evolution of articulatory control (brain and periphery) in the hominid line. My claim is that the control of articulation become more

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

sophisticated because protosign established a selective advantage in "lifting" the prior flexibility of manual control to vocal control *because of the drive to express novel meanings*. Further, **Aboitiz** and **FC&F** have little to say about language (lexicon, grammar, etc.) as distinct from articulatory control for speech sounds, whereas MSH explains how humans came to have compositional semantics (H-Chap.10).

Aboitiz asserts there are no convincing evolutionary relicts of the protosign stage during normal human development. However, "... even the hearing child makes extensive use of gesture in the transition to speech ... Iverson et al. (1994) found that gestures were more prevalent than (vocal) words in the children they observed as 16 month olds, whereas the majority of children had more words than gestures by 20 months of age. (H298)." See also comments by **Emmorey** and **Diessel.**

Aboitiz notes some limited aspects of vocal articulation which display similarities between humans and monkeys and then notes that "this does not preclude the co-option of additional systems or motor programs like ingestive behaviors, or hand manipulation, to subserve the emerging new vocal system," citing Ferrari et al. (2003). This report demonstrates mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey, but does not support the idea that vocal articulation is primary. Moreover, H230–231 suggests that Ferrari et al.'s claim that "Ingestive actions are the basis on which communication is built" might better be reduced to "Ingestive actions are the basis on which communication about feeding is built" (a path to oral pantomime?).

Pazzaglia has found that lateralized frontal structures involved in the execution of actions are also involved in the discrimination of non-linguistic sounds produced by buccofacial actions, such as a kissing sound, and with sound recognition related to limb actions. However, the data relate functions to large brain structures, losing the fine discrimination of circuits afforded by macaque neurophysiology. **Pazzaglia** asserts that macaque audiovisual mirror neurons (H133) should be given more emphasis in accounting for the evolution of spoken language, but the audio response of these neurons is to a manual, not a vocal, action. Most of our actions can be evoked by multimodal cues – we may, e.g. see, hear or smell a predator but react in the same way – and this auditory relevance to action does not link to communication in the way Fogassi and Ferrari's oro-facial neurons do, at least for feeding. So the issue is whether we should emphasize auditory mirror neurons for vocal actions at an early stage of the posited evolution, or expect them to "blossom" only after early protosign is able to scaffold protospeech. In any case, we may agree that human evolution exploits "ancient" auditory pathways but expands auditory motor control and working memory (cf. **Aboitiz**) to bring in a new level of vocal control that – unlike birdsong – links to

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

semantics. As **Pazzaglia** notes, songbirds do have mirror-like neurons (Prather et al. 2008) but here we tap the general discussion (H241–244) of whether vocal learning in humans arose much as it did for songbirds, with semantics being the uniquely human sequel, or whether human evolution was "semantics first," with vocal learning building on prior skills in manual communication.

FC&F summarize their important study of conditioning whereby partial voluntary vocal control is available in macaque ventral premotor cortex, and so may have been available in LAC-m. However, this very limited control is not demonstrably relevant to monkey communication in the wild. A possible counter-argument is that matching a signifier with a signified (in neural code) is just a form of conditioning. The fact remains that monkeys can assemble complex manual behaviors but not complex vocal behaviors. **FC&F** do note the poor level of vocal control reached by the macaques (very different, indeed, from the babbling of human infants). Even after extensive conditioning "about half of the trials consisted in failed attempts to vocalize in which the articulatory oro-facial gestures involved in the coos were made without sound emission." The connections revealed by their study reflect massive conditioning of monkeys in the laboratory – brains have many connections which did not evolve for a specific function but rather "failed to be pruned" and yet can be exploited by massive training. For me this militates against the view that protospeech did not need scaffolding from protosign, but does indicate pathways presumably present in LAC-m whose massive strengthening in post LAC-c human evolution could have come to support articulatory control adequate for protospeech.

FC&F note imaging studies in chimpanzees demonstrating activation of the homolog of human Broca's area during the production of communicative vocal and hand gestures and that the same region is also involved in communicative oro-facial/vocal signaling. They find these data "hard to reconcile with an origin of language based only on the brachio-manual gestural communication system." For them, some evolutionary pressure for extended vocal control must have come into play *well before* the use of protosign developed. I replace "well before" by "once." I am not saying "no prior vocalization before protosign," but they provide no counter-argument to my claim that it was pantomime-based protosign that opened up semantics.

In many papers, **Aboitiz** has emphasized the expansion of working memory (WM) in the evolution of brain mechanisms supporting language, stressing Baddeley's (e.g. 2007) "phonological loop." This is a valuable complement to MSH, but note that Baddeley's model also include a "visuospatial sketchpad." Of particular relevance, then, is my group's development of Template Construction Grammar to model the description of visual scenes (H-Fig.2–12; Barrès and Lee, 2013).

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

For **Aboitiz**, the amplification of a dorsal auditory pathway to the ventrolateral prefrontal cortex was a key event in evolution of the human brain: "*As the ability of vocal learning developed*, the primitive dorsal auditory-vocal circuit started to expand, recruiting neighboring circuits involved in other aspects like gestures and also hand control. Specifically, the inferior parietal lobule was recruited to *support working memory for vocalizations*, by specifying motor goals based on sensory (acoustic) information. [My italics]" But what was the evolutionary pressure for this development of vocal learning and working memory? **Aboitiz** does not tell us. Perhaps he agrees with those who argue that humans evolved "meaningless" song akin to birdsong first, and that this was infused with meaning only at a later stage ("phonology first, meaning second," H243) but he offers no argument to refute my hypothesis that it was the scaffolding of meaningful protosign that provided the evolutionary setting for increased articulatory control in the human line.

6 Homologies between macaque and human brains

The enterprise of which MSH is a part includes two complementary efforts: To place the language-readiness of the human brain in evolutionary perspective, and to offer an enriched account of neurolinguistics (the search to understand the human brain's mechanisms for supporting language). Clearly, each effort can inform the other. The study of homologies between macaque and human brains is an important part of that effort (Arbib and Bota 2003), and several commentaries usefully augment the sketch offered in H-Chaps.4and5. Here, I simply note that Hecht et al. (2012) link species differences in mirror system connectivity and responsivity in macaque, chimpanzee and human with species differences in imitation and social learning of tool use.

7 Action, hierarchy and meaning

FC&F note that mirror neurons can generalize their response to motor acts that, after training, were incorporated in the monkey motor repertoire (as indeed I discuss, and model computationally, H128–135) and that a percentage of mirror neurons discharged during observation of tool use actions performed by the experimenter which were not in the monkey's repertoire. (So the latter are not mirror neurons in *sensu stricto*; cf. my discussion of *potential mirror neurons* and *quasi-mirror neurons*, H132). They conclude that the motor system can "extend

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

the capacity of understanding goals to observed actions that have not been motorically experienced." However (H141), it is a mistake to argue from activity in the mirror system that "understanding" is mediated by that system rather than by a larger system of which it may (but need not always) be part. **FC&F** assert that if, as I propose, a mirror system for pantomime (as precondition for protosign) emerged, it is plausible that the neural plasticity they describe represents a substrate from which a system matching action observation with action execution may have expanded, incorporating also several types of intransitive gestures endowed with new meanings. Indeed, this is what I said (H132) for pantomime, but I reiterate that it is unlikely that the understanding of the meaning of pantomime or protosign or language rests within the mirror system alone. Rather (Figure 1.Left), I argue that a dorsal mirror system supports the recognition of *articulatory form*, but that understanding of the *meanings* of words rests on a complementary ventral system.

Importantly, **FC&F** emphasize the capacity of the cortical motor system to organize action sequences and the need of vocal and gestural communication to combine and control different effectors in order to produce complex social signals. I agree. **FC&F** assess important studies from their lab of the responses of parietal (area PFG) and premotor (area F5) grasping neurons during execution and observation of natural action sequences. Monkeys were trained to place a piece of food in a container if the container were present, but were otherwise free to eat the food. They observed activity of neurons during the initial phase of reaching to grasp the food. Some neurons were active during this phase only in the container condition, others in the reach-to-eat condition. Notably, this differential response was shown also by mirror neurons during observation of grasping by another individual. Chersi et al. (2011) conclude that neurons in parietal and premotor cortex are organized in motor chains, each coding a specific action goal, but their data do not support **FC&F**'s claim that "the premotor-parietal motor system plus the prefrontal cortex can provide a substrate for . . . *hierarchical* combination of motor elements. [My italics]" Two-element chains give us no insight into how the hierarchical structuring of novel sequential actions (praxic or communicative) is produced and recognized in the brain. In particular, we need to distinguish the coordination of articulators in a single word (or call) from the flexible sequencing of words based on different hierarchical structures in different contexts Indeed, it seems likely that connections with the basal ganglia play an essential role (H-Figure 4–14). Note the opportunistic scheduling of the augmented competitive queuing model (H139) and the extension of a model of sequence learning via interactions of basal ganglia and cerebral cortex to model a simple version of using constructions to extract the semantics of a sequence of words (Dominey et al. 2006). Much further modeling is required.

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

FC&F suggest that the human inferior frontal gyrus is part of the mirror neuron system, but I would rather say it is "associated with" this system since it contains far more than mirror neurons and receives input by multiple pathways. They cite evidence that Broca's aphasia (which may involve lesions to areas other than Broca's region) may involve impairments not only in phono-articulation but also, depending on the extent of the lesion and on the involvement of the nearby areas, in the processing of the hierarchical structure of a sentence But such generic observations do not tell us what pathways and related systems are impaired – and a fortiori omit key elements of the evolutionary pathway from the macaque mirror system to the human language system (but see Fogassi and Ferrari 2012 for further discussion).

Tettamanti claims that motor and cognitive behavior involve "linear sequential hierarchies [that] are transparent to the senses and can in principle be coded by the [mirror system]." Motor behavior in general involves far more hierarchical structure than the chains of Chersi et al. However, we can certainly agree about the importance of brain regions beyond the mirror systems in syntactic computation and integration – though I would emphasize (as in construction grammar) that grammatical processes rarely invoke syntax without integrating semantic cues (but see Barrès and Lee 2013 for further subtleties. Interestingly, **Tettamanti** suggests that the distinction between local and non-local syntactic dependencies may largely reflect the type of syntactic information that can be extracted in early child language acquisition through generalization from lexically specific constructions and, later in development, the maturation of non-perceptual, internal cognitive skills related to hierarchical structuring – compatible with usage-based approaches such as construction grammar.

Different languages that are far apart in their historical and geographical origins may share common features that may have originated independently several times, such as the basic subject-object-verb order or rare grammatical properties such as evidentiality. **Tettamanti** thus suggests that complex language structure is not solely the product of cultural evolution, but is intrinsically shaped by the manner in which our brain is perceptually and computationally constrained in elaborating the physical world. I would rather say that complex language structure is the product of cultural evolution, which is intrinsically shaped by the manner in which our brain is perceptually and computationally constrained in elaborating the physical *and social* world. **Tettamanti** tackles the daunting task, still beyond our reach, of assessing how brain imaging might tease apart biological versus cultural evolution of language. He asserts that neuroimaging of multilingual speakers has consistently demonstrated that multiple spoken languages are largely represented in overlapping left-hemispheric perisylvian networks and concludes that "the highly conserved syntactic patterns across languages are the

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

product of biological, rather than cultural, evolution." This seems to me unjustified. Consider similarities in neural correlates for people reading diverse phonological scripts. We know these are the product of development channeled by similar cultural innovations. Against this we may place the distinct neural correlates for reading Chinese characters. For example, Chen et al. (2002) use fMRI to show a direct contrast between the neural correlates of reading Chinese via characters versus the pinyin alphabet. Moreover, the studies reported by **Tettamanti** are coarse grain, inadequate to tease apart the specifics of different languages. By contrast, Allen et al. (2012) used multi-voxel pattern analysis with fMRI to distinguish neural correlates of closely related grammatical constructions such as the dative (e.g. Sally gave a book to Joe) and the ditransitive (e.g. Sally gave Joe a book). Combination of areas BA22 and BA47 were sufficient to distinguish the two constructions better than the controls and better than chance but it seems improbable that the distinction between these constructions is biologically evolved whatever the relevance of the two areas, both implicated in semantics, may be.

I agree with **Tettamanti** that "[f]uture neuroimaging studies will need to boost longitudinal research in newborns and young children with the hope to resolve the neural basis of incremental syntactic structure formation" – but I stress that we must avoid a bias towards autonomous syntax rather than integrating syntax and semantics. **Tettamanti** also argues for hyperscanning to study multiple subjects, each in a separate magnetic resonance scanner, interacting with one another in ecological (from an evolutionary point of view) but strictly controlled, experimental paradigms to let participants invent full-fledged languages, through mutual interaction, akin to my cultural evolutionary scenario, but in a much more compressed time frame. Another approach would apply neuroimaging to subjects engaged in emerging sign languages or pidgin and creole languages. Leaving neuroimaging and syntax aside, we may note that Fay and Lim (2012) recreated a simplified historical record under laboratory conditions by having modern humans communicate a set of recurring concepts to a partner while prohibiting participants from using their existing language system. Participants played six games, using the same item set on each game (presented in a different random order). Communication accuracy increased across games in each condition. However, accuracy was higher at game 6 in the gesture (94.4%) and vocalization plus gesture conditions (95.8%) when compared to vocalization only (52.1%): gesture is more effective than vocalization. Their findings, although compromised by using modem humans, lend support to the view that spoken language arose out of manual gestures (Fay, Arbib, & Garrod, 2013).

Dominey advocates the construction grammar framework but asserts that I fail to address the generative capability of human language. This claim seems mistaken, but I agree that it is a worthwhile exercise to assess the relation between

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

my group's template construction grammar (H65–71) and work on embodied construction grammar, his own work and, one might add, fluid construction grammar (Steels 2003). Unfortunately, **Dominey** gives no details of his own ingenious use of computer models and human-robot interaction as a testbed to address issues related to compositionality in semantics and syntax. He stresses Levinson's observations on the mechanisms of mental inference needed for turn taking in a conversation. Levinson argues that before language evolved, there was already an "interaction engine" that allowed individuals to understand communicative actions not just as behaviors but as intentional communicative actions. To build on this, note Menenti et al.'s (2012) work toward a neural basis of interactive alignment in conversation, work of Pickering and Garrod discussed later, Jeannerod's (2005) linkage of mind reading to mirror system parity mechanisms and Levinson's (2013) extension of his discussion to the group interactions involved in music making. My own group has just begun dyadic modeling (modeling how the brains of two agents perform and change as the agents interact) of the emergence of chimpanzee gestures (Arbib et al. 2013, Gasser et al. 2013).

Dominey posits that the combinatorial structure of language arose to transmit messages constructed from an equally combinatorial system of thoughts, with the precedence for combinatoriality lying in the conceptual system. The holophrasis argument (H-Chap.10) is that the interaction between a scene and a holophrase can via fractionation move us toward grammar. But I would further argue that the perceptual system is not previously "carved up" in a way that lets this compositionality be picked off. As I noted earlier, I suggest that the combinatorial conceptual system emerges only as language emerges. One cannot evolve it *a priori* before there is at least the rudiment of protolanguage to which it can serve as an interface.

Dominey reviews a number of papers in seeking a neurophysiology of meaning, but his argument is too connectomic to get at the changes in processing capability that expanded and strengthened the network he posits. Here my discussion (Arbib in press) of Vigliocco et al. (2011) on the neural correlates of actions and objects versus nouns and verbs may be relevant. But the challenge of developing a neurolinguistics (let alone an evolutionary hypothesis) for successive abstractions, function words, tense, counterfactuals and more is very much open.

8 The mystery of motivation

Zuberbühler stresses the cooperative motivation that characterizes human communication. Unlike other primates, humans routinely base acts of communication on assumptions and knowledge that they share with receivers. The ability

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

to take into account what information is novel and interesting for a receiver develops early in human infants, and may therefore not require advanced theory of mind abilities (Liebal et al. 2010). There are currently insufficient data to decide how important the ability to take others into account is in non-human primate communication. **Zuberbühler** suggests that a key event for human evolution was the migration out of the forested habitat, the home of most non-human primates including all great apes, into the savannah. (By looking at sediment core evidence of plant leaf waxes and pollen blown from the land into the Gulf of Aden, Feakins et al. (2013) determined that the savannah developed 12 million years ago, a full 6 million years or more before human ancestors became bipedal.) Survival and reproduction in the open savannah may have been more challenging for earlier humans due to predation, intergroup conflicts, and new demands in terms of cooperative breeding and foraging and this may have fostered hyper-sociality and hyper-cooperativity. Equipped with an already very efficient communication device and high social skills, perhaps similar to what is seen in chimpanzees and bonobos, early humans were well positioned to evolve more efficient communication.

As **Dominey** notes, we could explain how a system could observe a scene and then describe it without any insight into why the system would be motivated to produce such an utterance. "*Why* did the brain get language?" An account of the evolution of the language-ready brain must address the human motivation to represent and share the psychological states of others. **Dominey** notes research by Tomasello and his colleagues who compared apes and human infants to show that humans display a unique motivation to share mental states that is present "before the emergence of language." (See my discussion of "A Cooperative Framework," H198, and note **Sinha's** comments on niche construction.) Indeed, intended communication is the second of the properties listed (H164) as necessary to establish protolanguage. **Dominey** cites Syal and Finlay (2011) for their suggestion that linking strong motivational inputs to circuits related to perceptual and motor aspects of communicative signaling (including vocalization, gaze and facial expression) could have contributed to the development of socially motivated communication but this seems more a rephrasing of the problem than a solution. More specifically, they review evidence of development of vocal communication across species, particularly birdsong, and new research on the neural organization and evolution of social and motivational circuitry to suggest that human language is the result of an obligatory link of a powerful cortico-striatal learning system, and subcortical socio-motivational circuitry. (See the earlier discussion of basal ganglia, including **Dominey**'s own contributions to its modeling.) Note that the medial call system in primates is strongly linked to emotional states, and that MSH outlines both how humans developed lateral systems to gain the open

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

semantics that birdsong lacks and built up links between the lateral and medial system. This may be the basis for developing a fuller response to **Dominey**'s challenge. Deacon's (1998, 2006, 1997) work, linking semiotics to brain research, draws attention to changes in the structure of the brain's motivational systems as essential to making the interpretative function of the human brain possible, a dimension underplayed in MSH.

For **Dominey**, *human language cannot originate without a self*. He usefully contrasts my notion of a self as a collection of social schemas (actually, internal schemas which may in part reflect accommodation of social schemas, H18) with Neisser's four developmental stages of self: the ecological, interpersonal, conceptual and, finally, temporally extended self. "[T]here is no reason why [Arbib's] schema theory could not accommodate Neisser's levels of self. But this should be made explicit, because . . . it is perhaps the influence of the resulting 'social schemas' that will be part of the underlying drive to communicate. . . . Arbib's formulation of social schemas as collective patterns of behavior in a society does not appear to include [the] motivation to share mental states at the individual level." Indeed, my theory takes it for granted that humans want to interiorize (or rebel against) social schemas, but does not explicitly address what motivates them to do so. A hook for remedying this gap is given by my brief discussion assessing "how our ability to experience others' emotions affected the way the language-ready brain has evolved. (H144)" Relating this to my Property 2 (H164), the challenge is to tease out the motivational mechanisms whereby communication became *intended* communication. Presumably, this can best be understood in terms of social interaction in general, rather than being specific to symbolic communication.

The switch to "intentional communication" is baldly stated in H-Chap.6, but the working out of how the underlying brain mechanisms evolved is left untouched. While **Dominey** (with Tomasello) offers another contrast comparable to my simple versus complex imitation, he offers no help as to how to fill in the neuro-details beyond citing Syal and Finlay (2011). I have some work on emotion (Arbib 2005; Arbib and Fellous 2004) but have not integrated it into my account of language evolution. We must take into account the ancient roots of social interaction (e.g. Edwards and Spitzer 2006: show that the relation between serotonin and social dominance can be seen even in our very distant relative the crayfish) so there is much to build on in seeing how systems of social motivation became uniquely human.

Dominey suggests that the notion that language is there to allow the infant and caregiver to socially engage is missing, but see my discussion of *assisted imitation* where the caregiver uses language to motivate the child even when the child does not understand it (H198–201). Zukow-Goldring (2012) goes further,

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

showing how assisted imitation may seed language development. Where Meltzoff (2005) builds his account of children's cognitive development on the hypothesis that "the other is like me," Zukow-Goldring emphasizes the complementary "I am like the other" which is necessary for the child to gain skills outside her original repertoire. Clearly, the relevant comparative primatology must assess the claim (Penn and Povinelli 2007) that there is no evidence that non-human animals possess anything remotely resembling a "Theory of Mind." However, I do take seriously that the issue of representing others in intention-based interaction deserves more attention than I give it. Interestingly, it has been suggested that an important ground for recursion is that language is particularly well-suited to human thinking about propositional attitudes: understanding others' desires, beliefs, knowledge states, emotions and so forth (de Villiers and de Villiers 2009), an important challenge for future work.

9 Learning from apraxia

Cossu et al. (2012) compared high functioning ASD children with typically developing children on three domains of motor cognition: 1) imitation of actions, 2) production of pantomimes, and 3) comprehension of pantomimes. ASD children fared significantly worse in each domain. In comprehension of pantomimes, ASD children came close to the younger typically developing children's level of performance; yet fared significantly worse with respect to age-matched controls. Overall, ASD children reveal a profound damage to the mechanisms that control both production and pre-cognitive "comprehension" of the motor representation of actions. They suggest that many of the social cognitive impairments associated with ASD are rooted in their shortcomings in the domains of goal-related motor behavior and speculate that this might be partly due to early problems with the mirror neuron system (Arbib 2007; Williams 2008; Williams et al. 2001). More generally, Gallese et al. (2009) assess the role of motor cognition in the phylogeny and ontogeny of action understanding.

Pazzaglia reviews a range of apraxia studies, noting that although apraxia is not explained by [or explanatory of] defects in language, a common neural substrate of the left brain is shared by patients who exhibit impairment in the linguistic and action domains. Her recent studies show that the lesions of left inferior frontal gyrus are involved not only in deficits of gestural execution and imitation but also with deficits in object-related (transitive) and non-object-related (intransitive) gesture discrimination. The evolutionary transition from transitive actions on objects (praxis) to intransitive pantomime (communication) requires fuller explanation in the development of MSH. The relation between apraxia and aphasia

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

has been a central concern (Arbib 2006). In brief, the argument is that lesions underlying aphasia and apraxia overlap but that distinct circuits with a partially shared evolutionary history serve language and praxis in the modern human brain (recall Figure 1.Left). Evolution of brain regions may involve parcellation into classes of modules, the progressive fusion of modules of like class, and the eventual separation of fused modules into two or more areas. Reduplication of circuitry may form the basis for differential evolution of the copies to serve a variety of functions. Hence data on differential localization of limb praxis, speech and language, and emotional communication may offer insight into what form these patterns of reduplication and subsequent differentiation might have taken and thus fill in details currently glossed over by MSH.

Pazzaglia reports studies of primary progressive aphasia in which gesture-discrimination and gesture-imitation deficits were closely associated with impairments in verbal tasks that require the translation of a visual input into a motor output, such as writing to dictation and the repetition of words and pseudo words. One wonders what defects would show up in the scene description task we modeled with Template Construction Grammar. She cautions that although the implicated brain areas are anteriorly connected to the frontal operculum, the language and praxic impairments have only been partly mapped to regions considered crucial nodes of the human mirror system. Such uncertainties highlight the complementary challenges of developing an evolutionary model and a model of the modern functioning human brain – and the attempt to render them consistent.

Pazzaglia suggests that perceptual motor codes potentially play a role in the motor representation of action/word outcomes and might critically have an impact on action/word comprehension, but we must distinguish priming of, e.g., a foot or hand action by a verb from the more general case of the perception and production of the articulatory form of a word whose semantics cannot be related to a motor area. (And, of course, knowing that something is a hand action is only a small part of its meaning – e.g. *hit* versus *caress*.) She finds it likely that the human brain builds and plastically adapts distributed representations, using a process of simulation in the sensorimotor brain. However, I find this view too reductive, and have argued that the brain needs new mechanisms beyond the generic primate sensorimotor system to support metaphor and abstraction, let alone the constructions of a modern human language (Arbib 2008, H271–278), without denying that the child gets started within a language community on the basis of her embodied experience.

Just as **Dominey** emphasizes Levinson's concerns with the unfolding of language in conversation, so **Pazzaglia** cites Pickering and Garrod's (2013) integrated theory of language production and comprehension to suggest that speakers construct forward models of their linguistic/motor act before producing it,

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

whereas perceivers covertly imitate those acts and then construct forward models of the same act to predict and monitor the upcoming utterances. Indeed, the first paper on MSH (Arbib and Rizzolatti 1997) discussed the role of forward and inverse models; and a sequel (Oztop et al. 2005) to the MNS model of learning in the mirror system for grasping (H129) provides a computational model that builds upon visuo-manual feedback control to implement mental simulation and mental state inference: control mechanisms for manipulation are readily endowed with visual and predictive processing capabilities, allowing a natural extension to the understanding of movements performed by others. One may see in this the precursor for prediction and mental state inference in conversation. Skipper et al. (2007) assess the role of pairs of inverse and forward models in coupling auditory and visual aspects of listening. Nonetheless, in much of our listening, we must expect the unexpected.

Pazzaglia cites data supporting the notion of "a matching mechanism in which auditory stimuli directly and specifically retrieve an invariant motor representation." Reduced word comprehension was observed in children with severe neurological conditions yielding motor deficits that affected articulation when compared with individuals with comparable motor deficits that did not involve language articulation. This suggests that the phonological matching of speech perception needs to be evoked for early language development. However, **Stasenko, Garcea, Mahon (SG&M)** oppose the motor theory of speech perception implied here. They ask what happens to Liberman's motor theory of speech perception when the motor system is damaged. Although the 1998 version of MSH espoused the motor theory, the phrase "motor theory" appears nowhere in the book. I do write (H37) that the success of language is based on the *parity principle* (Liberman and Mattingly 1989) – the fact that, more often than not, the meaning understood by the receiver is (at least approximately) the meaning intended by the sender. **SG&M** note, correctly, that my "core argument about the evolution of language . . . could be true even if strong forms of the motor theory of speech perception are false." However, since many people equate MSH with the motor theory, **SG&M**'s critique is of value in the present context. The book's MSH posits that parity is at the level of the communicative gesture (e.g., word as articulatory unit) – but, since we can recognize words even when pronounced with an accent very different from our own, recognition of meaning requires a path from recognition of auditory word form to the ventral system (Fig.1.Left) rather than, in general, mapping of auditory form to motor gesture, unless our aim is to imitate the form we hear. I suggest that "the understanding of *all* actions involves general mechanisms that need not involve the mirror system strongly – but that for actions that are in the observer's repertoire, these general mechanisms may be complemented by activity in the mirror system that enriches that understanding by

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

access to a network of associations linked to the observer's own performance of such actions [H 142]," and are crucial for acquisition and imitation.

SG&M note the recent [sic] and widespread interest in how cognitive and perceptual processes are inherently *active* processes; I cannot resist noting my "unrecent" stress on action-oriented perception (Arbib 1972). They ask: "If recognition can be understood as an active process whereby implicit 'hypotheses' are generated about the nature of the percept, then what is the format of the information over which those 'hypotheses' are formulated?" **SG&M** observe that we lack a theory of the dynamics of activation spread between input and output systems and how that spreading of activation may be gated. I would agree, but suggest that "spread of activation" is too restrictive a view, and argue (Arbib in press) for the importance of the competition and cooperation of schemas (H10–18). Further, **SG&M** review cases in which motor processes can be impaired (at multiple levels) while recognition processes remain intact (again, at multiple levels). Indeed, **Pazzaglia** warns that, despite the richness of data she reviews showing convergence of the neural and functional aspects linked to actions and linguistic processes, one must note the heterogeneity of symptoms across clinical populations such that dissociations occur for damage at any level of the input–output patterns. She cautions that neuropsychological deficits rarely result in the total loss of a given ability so the association in two disturbances may denote a simple link between language information and motor activity, or may reflect a unitary mechanism of direct matching subserving input-output systems (see the earlier discussion of apraxia with aphasia).

SG&M recall that chinchillas can discriminate human speech to indicate that a production system is not necessary for recognition to occur. In the cited study (Kuhl and Miller 1975), four chinchillas were trained to respond differentially to syllables with /t/ and /d/ consonants produced by four speakers in three vowel contexts. However, discriminating a couple of phonemes is not the same as disentangling a fast verbal performance into a hierarchical structure of words to extract the meaning of the sentence. Auditory discrimination evolved long before language. Indeed, dogs may have an extensive recognition vocabulary for human words (H291–293) but have no capacity for grammar or production.

Noting that infants discriminate sounds that they cannot yet produce, **SG&M** doubt that there is any coherent developmental story to account for why the production system would be used for recognition – but see the discussion of *potential mirror neurons* and *quasi-mirror neurons* (H132) which may have the potential to become mirror neurons through learning, but manifest only the motor or perceptual correlates of the actions with which they are associated (perhaps as part of a population code). The MNS models (H128–135) show how an infant who has already acquired an action may, through observation of self-performance, learn

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

to recognize viewer-independent trajectories as performed by others. Imitation works in the opposite direction (H190): Visual recognition of how another achieves a goal must come first as a basis for feedback in acquiring motor control. Alas, none of the commentators discussed the case I make for the value of computational modeling – not in the sense of reducing the brain to current digital computers, but rather in seeking descriptions of cooperative computation explicit enough to be *simulated* on a modern electronic computer (H4–16). I thus draw attention to the (all too brief) "Notes Towards Related Modeling" (H210–212) at the end of the chapter on imitation.

SG&M consider data indicating that TMS to motor areas modulates motor-evoked potentials (MEPs) recorded from the tongue while listening to speech sounds whose production engages tongue movement. **SG&M** favor an explanation for such data in which activation spreads from perceptual levels of processing through to motor processes. I favor a priming view for this effect – it will be of little importance in "easy" recognition, but may play an important role in imitation and in "hard" recognition.

As **SG&M** note, the original claim for the motor theory of speech perception was that using the production system "solved" the invariance problem and then raised the problem "that in order to match the input to a gesture representation, the acoustic information must be parsed – which would seem to introduce some circularity into the motor theory of speech perception." Certainly, the MNS models (H128–135) require perception to activate the mirror neurons and explain how perception generalized across multiple trajectories arises through learning. There is no automatic "resonance," to use a term overly popular in the mirror neuron literature. In related work, Moulin-Frier and Arbib (2013) model how a listener comes to understand the speech of someone speaking the listener's native language with a foreign accent. The core idea is that the listener uses (implicitly) hypotheses about the word the speaker is currently uttering to update probabilities linking the sound produced by the speaker to phonemes in the native language repertoire of the listener, thus speeding the recognition of later words, on average. This task seems to fly in the face of the motor theory, and the paper thus assesses claims for and against it and the relevance of mirror neurons.

SG&M cite Pazzaglia's finding that patients with buccofacial apraxia had differential impairments for recognizing mouth compared to hand-generated sounds, while patients with limb apraxia had differential impairments for recognizing hand-generated compared to mouth-generated sounds, but then note that although limb apraxics cannot use tools skillfully, their ability to recognize actions and objects can remain intact. Here, I add Aziz-Zadeh et al.'s (2012) observations of a woman who, born without limbs, engages her own sensory-motor representations as a means to understand other people's body actions. They argue

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

that, when observed actions are not possible, mentalizing mechanisms, relying on neural structures outside the mirror system, are additionally recruited to process the actions. As in my assessment of Buccino et al. (2004), I suggest that non-motor pathways are engaged whether or not an action is in one's repertoire. However, given our earlier discussion, it seems that if one habitually observes certain actions of others, one may form a forward model to predict those actions even if they are not in one's repertoire – but such models will perforce be less detailed in the latter case. Indeed, professional basketball players predict the success of free shots at a basket earlier and more accurately than sports journalists and novices (Aglioti et al. 2008). Moreover, performance differed even before the ball was seen to leave the hands. Aglioti et al. suggest that "fine-tuning of specific anticipatory 'resonance' mechanisms . . . endow elite athletes' brains with the ability to predict others' actions ahead of their realization," but (again) I prefer to avoid the term "resonance" for a detailed forward model developed through both sensory and (usually) motor experience.

SG&M report that deactivation of the entire left hemisphere usually results in a complete failure to produce speech but phonemic discrimination error rates may remain below 10% as shown by an auditory word-to-picture matching task. This seems to demonstrate a right hemisphere ability to recognize concrete words without ruling out a right hemisphere mirror system. It certainly raises the challenge of understanding when the right hemisphere may have some capability for left hemisphere functions; and these may provide a platform for compensatory changes following stroke or lesions.

Some patients with Broca's aphasia have deficits in speech sound recognition, others do not. This raises the stakes on what Broca's area does, noting that lesions to Broca's area alone seem insufficient to yield agrammatism, and lesions excluding Broca's area may yield it. Agrammatism data suggests that word production may be relatively spared but invocation of constructions is not. **SG&M** cite a study of patients with lesions to the "human mirror system" on tests of word comprehension and syllable discrimination in which the worst performance on comprehension tasks was in patients with damage to the temporal lobe and see this as consistent with the view that the temporal lobes are necessary and sufficient for speech perception. Barrès and Lee (2013) offer a preliminary modeling perspective but the database on neural correlates of actual processing is far from complete. I believe that learning what is really going on will not be settled without detailed computational modeling.

Noting that my position (Fig.1.Left) is largely in line with the Hickok-Poeppel proposal of ventral and dorsal paths for speech processing, **SG&M** see the critical issue for MSH to be whether my framework assigns a causal (i.e. necessary) role to motor systems in parsing the speech signal. (I would note that a cooperative

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

computation approach addresses *multiple* causes – so that any process may be causal yet rarely necessary.) They conclude that my theory is largely independent of motor theories of perception. They do concede that the motor system is highly interconnected with, and relevant in some as yet unspecified way, to perception. The question then becomes whether (as I hold) motor information can facilitate recognition of speech under degraded conditions, or provide top-down constraints that may assist in guiding the formulation of hypotheses over auditory information. "Analysis by synthesis" can occur in the auditory domain, and can be informed by relevant information that is represented and processed by other systems (including, but not limited to the motor system). "This suggests a shift in research, from demonstrations of the 'mere fact' that the motor system is activated during perception, to research aimed at unpacking the processing dynamics that mediate interactions between input and output systems." And that's my point.

10 Learning from linguistics: the power of deixis

Diessel advances MSH by arguing that demonstratives "constitute a unique class of expressions that speakers of all languages use in combination with pointing gestures to establish joint attention, a cognitive . . . [process essential for] imitation. No other linguistic device is so closely associated with the body and gesture . . . [and] demonstratives are not only used to direct the interlocutors' attention to concrete entities in the outside world, . . . but also to organize the information flow in discourse, which in turn leads to their development into grammatical markers. In this way, demonstratives provide an explicit link between gesture, imitation, and grammar that is consistent with Arbib's theory of language evolution." These ideas are relevant to discussion of pointing being limited in captive apes and apparently absent in apes in the wild (H81–82) and to the language acquisition chapter, where more attention could have been given to the crucial early role of gesture in general and pointing in particular.

As **Diessel** observes, the gestural use of demonstratives provides a powerful mechanism for the child to engage in verbal activities with a limited vocabulary. With age, language becomes more independent from gesture and situational cues though demonstratives continue to play an important role even in adult language.

Diessel's discussion of demonstratives and the emergence of grammar is clearly related to the grammaticalization section (H335–344) but strengthens its relevance to MSH (and to the ape and acquisition chapters) by his argument that demonstratives have to be kept separate from both content words and function words.

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

11 When did language evolve in the human lineage?

Sinha agrees with me that (a) the evolution of the "language-ready brain" preceded the development of language "proper" (what he calls *evolutionary modern languag*e, EML, and what I call language *simpliciter*); and (b) EML is a relatively late human acquisition or artifact. However, **Dubreuil and Henshilwood** disagree.

I presented eleven properties that make the use of language possible (H62), claiming that the last four required no new brain mechanisms but emerged through the operation of cultural evolution on language-ready brains of *Homo sapiens*. But **Sinha** seems to misread me as claiming that there was no cultural evolution before the divide. Not so. I noted that, for example, chimpanzees have distinct "cultures" in different parts of Africa (H156). Moreover, the emergence of pantomime and protosign reflected the emergence of brain structures made advantageous through cultural evolution. However, I agree with **Sinha** that "What makes humans unique is . . . not inscribed [only] in the human genome . . . [T]he human biocultural complex . . . constitutes a self-made environment for adaptive selection." I add that Atsushi Iriki has recently advanced the notion of *triadic (ecological, neural, cognitive) niche construction* (Iriki and Taoka 2012); Arbib and Iriki (2013) offer a partial integration of this with MSH. My thesis is that bio-cultural evolution supporting protolanguage coupled with a language-ready brain (and, more generally, a brain ready for rich enculturation: Whitehead 2010) laid the basis for further non-biological cultural evolution to finally yield languages with the expanded lexicon and grammar missing in protolanguages. **Sinha**'s valid concern with the social milieu that protolanguage made possible – the constructed niche in which language could (culturally) evolve – could be tied to **Dominey**'s concern with *why* we want to talk: the motivation to communicate using protolanguage would be part of the bio-cultural niche from which *H. sapiens* developed language (thus pushing the question back closer to LAC-c).

Sinha says that I do not make it clear what I mean by "real" language, or what would be the simplest form of real language. Indeed, I explicitly refuse to characterize the "simplest form of real language," emphasizing instead the notion of a *protolanguage spectrum* from very simple protolanguages to complex protolanguages which might also be viewed as simple languages (H253). He suggests a three stage division

Protolanguages → early languages → evolutionary modern languages

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

which seems to match my

Protolanguages → a spectrum of increasing complexity → languages

which replaces hard boundaries (as if there were a categorical distinction between protolanguage and "early language") by a spectrum of incremental changes as languages emerged from the simpler protolanguages.

Sinha adds: "it would be unwise to rule out the possibility that some genetic adaptations for language learning occurred after the speciation of *Homo sapiens*, during the long period that early languages were spoken." I have no trouble with that. Perhaps Baldwinian evolution (H156) was the means. Improvement in eye-hand-voice coordination, for example, could have improved and supported both praxic and communicative skills or, with **Aboitiz**, we could emphasize improvements in working memory. (Here I would like to acknowledge my debt to C.H. Waddington whom I met in the early 1960s and who introduced me to epigenetics (Waddington 1957); he also wrote on the "Baldwin effect" (Waddington 1953).)

Sinha emphasizes that *symbolic artifacts* (akin to my *social schemas*, H23–27), including writing and calendars, "are key components of human cognitive evolution, in virtue of their status as external representations of cultural and symbolic practices (Donald 1991), and embodiments of the 'ratchet effect' (Tomasello 1999) in cultural evolution" (compare H322–3). He uses his research on the language of time in Amondawa to question the often-assumed universality of the space-to-time metaphoric mapping and of a timeline independent of event-based time intervals. His analysis has echoes of my discussion of Nicaraguan Sign Language where, countering the view that the language emerged *de novo*, I note "the huge cultural input that goes into something like having a sign for each day of the week – a long way from the more-or-less spontaneous gestures of home sign (H306)."

Sinha hypothesizes that *multimodal protolanguage* has a time depth of at least 2 million to 1.5 million years and was almost certainly possessed by *Homo erectus*; *early languages* emerged "as the first original biocultural semiotic artifact of the language-ready brain" at 200 kya to 150 kya; while *evolutionary modern languages* probably date from 100 kya – 50 kya. Based on Stout's (2011) account of the evolution of stone tools, I identified four stages in the evolution of language (H324–330):

1. The simplicity of Oldowan industries suggests the absence of complex imitation skills so that communication among early *Homo habilis* and *Homo erectus* would have consisted of a limited repertoire of vocal and manual gestures akin to those of a group of modern great apes.

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

2. Acheulean industries are "transitional" between simple and complex imitation.
3. The late Acheulean and the emergence of *Homo sapiens* at about 200 kya marks the full emergence of complex imitation and the language-ready brain.
4. Rapid innovations in human material culture observed between 100 kya and 50 kya corresponds to the cultural evolution of fully-fledged human languages. It took tens of millennia for the language-ready brain to produce the types of languages with which we are now familiar.

Dubreuil and Henshilwood (**D&H**) bring their archeological expertise to bear on these claims. They concede that the production of Acheulean hand axes is arguably more demanding of working memory than that of Oldowan choppers and agree that Acheulean culture exhibited limited variation over almost one million years suggesting that "*Homo erectus* . . . had a limited capacity for innovation, at least in a technological sense." They ask whether this apparent limitation extended to language – though I would prefer to use "protolanguage" here. An argument against such a limitation might be that language developed for social interaction despite stasis of technological culture, with an account rooted in Theory of Mind, emotion and empathy, and social interaction as in "gossip theory" (Dunbar 1993) or inference from baboon social exchanges (Cheney and Seyfarth 2005; Seyfarth et al. 2005). For **D&H**, "Successfully adapting to a range of climates and environmental conditions suggests advanced levels of social cooperation and a technology that at least must have been adaptable." I would counter that protolanguage would suffice for the relevant social interactions. They do note a human-like life history, with prolonged infancy, would give human children more time for social learning while adults were engaged in cooperative rearing (Dubreuil 2010). This is my Property 7 (H166) which I associate with *proto*-language (and only later with language) and which, I suggest, required biological changes necessary for *H. sapiens*.

D&H suggest that "the brain of late *Homo erectus* was sufficiently 'language-ready' to permit the development of significantly complex syntactic structures, while high levels of cooperation [e.g. for hunting large animals] . . . indicate the presence of a strong motivation to communicate. It is unlikely that this motivation did not, during the hundreds of thousands of years of existence of those species, lead to significant grammaticalization." However, I see no need for "significantly complex syntactic structures" to coordinate a hunt. Bickerton (2009) emphasizes large-game scavenging rather than hunting as the niche in which language evolved, but I have argued that – whether for scavenging or hunting – protolanguage would suffice (Arbib 2011). Indeed, one may note that chimpanzees, too, engage in cooperative hunting and food sharing, but their prey, such as

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

colobus monkeys, are small (Stanford 1999), so we may certainly conclude that the lesser challenge of cooperatively hunting small prey does not even require protolanguage.

In assisted imitation (H199–201), when the caregiver assists a child to gain a skill, words – at least for the younger child – plays only a motivational role rather than invoking concepts already known to the child. Shaping a hand or guiding its trajectory can be done by actual guidance "in slow motion" as a basis for developing skill. If language already exists it may assist the process, but its role is secondary. And how important would syntax be for temporal planning for protohumans? Saying the protoword for "hunt-elephant" while vigorously waving a spear and then charging off in a particular direction might well suffice to recruit a team to go hunting, other protowords might indicate caution while approaching the prey, and yet others could provide basic cues which individuals could elaborate for themselves as the hunt proceeded.

The earliest known *Homo sapiens* fossils are dated about 195 kya and **D&H** list several innovations that appear between 100 and 60 kya while cautioning that several behavioral innovations appear significantly earlier, such as symbolic use of pigments ca. 200 kya and five fragments of pigment dated between 400 – 260 kya that show traces of use. "If pigment use is an archaeological indication of symbolic behavior . . . *and indirectly of language* [my italics], then the origin of these abilities, traditionally attributed to *Homo sapiens* has to be considered more ancient than commonly accepted." Fascinating data that, in important ways, take us beyond my brief discussion of stone tools. But note that sudden jump to "language." If I want to know what color shirt to wear, I can of course use language to consult my wife – but I can also just hold the shirt up against my body, and she can indicate by a shake of her head whether or not it is appropriate for that day. Thus, I do not see pigment use as requiring the use of language (or even "early language" in **Sinha**'s sense).

As **D&H** emphasize, well dated archeological sites between 300 kya and 100 kya are rare, so, although there might have been a sudden surge in human innovation at 100 kya, the possibility of a much longer and gradual evolution of modem behavior following a mosaic pattern is clearly probable. But this seems to me consistent with my view of language emerging across the protolanguage spectrum during the tens of millennia prior to 100 kya by bricolage, an emergentist view of language where special-purpose constructions (in the grammatical sense) are added to the available stock, but initially with limited generalization (H253).

As **D&H** note, "an essential attribute of cognitively modern societies is the capacity to create symbolic systems and to reflect these visibly in their material culture. Recognition of distinct symbols that impart different meanings by members of a social group requires that the material representations show

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

morphological variation that imparts to each an 'identity'." Here I think we are in the territory of **Sinha**'s symbolic artifacts. **D&H** argue that the innovative technologies and social practices observed in the archeological record, mostly at and after 100 kya in Africa, provide evidence that humans were capable of creating rich symbolic systems. They contend that "these innovations . . . are not best explained by the evolution of language strictly construed (whether biological or cultural), but by a change in social cognition and perspective taking . . . [T]hese people shared a symbolic system that had common meaning within their group and that this act of sharing codes applied probably both to their material and spiritual life." They further invoke "Theory of Mind." All this seems to me consistent with the bricolage view of language emergence: in some cases, symbolic artifacts emerged and spurred the creation of new linguistic forms to discuss them; in other cases, it was the emerging set of language tools that made new symbolic artifacts possible and catalyzed their dissemination.

For **D&H**, the brain was almost language-ready significantly before *Homo sapiens* and the cultural evolution of languages was well underway when the first *sapiens* evolved. "Limitations in perspective-taking and mind-reading abilities might have prevented some features of modern human languages from evolving, such as metalinguistic awareness, irony, and potentially some complex syntactical structures." Here I think we simply have to agree that the jury is out – neither **D&H** nor I have the data necessary to settle the issue. I cannot disprove their claim, but neither have they disproved my claim that only *Homo sapiens* had the neural means to progress beyond the early stages of the protolanguage spectrum. But I do want to reiterate the graded nature of the seven "biologically established" properties which I claim define the language-ready brain (H163–167).

Acknowledgments: I am most grateful to David Kemmerer for conceiving this Special Issue and bringing it to fruition. In addition I thank all the commentators for bringing their own expertise to bear. Carol Padden and So-One K. Hwang offered valuable insights into the rates of production and perception of sign language, Katja Liebal and Simone Pika commented on primate communication, and Giuseppe Luppino helped re-assess data on macaque neuroanatomy (though not all comments could be used in the space available).

References

Aglioti, S. M., P. Cesari, M. Romani & C. Urgesi. 2008. Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience* 11. 1109–1116.

Allen, K., F. Pereira, M. Botvinick & A. E. Goldberg. 2012. Distinguishing grammatical constructions with fMRI pattern analysis. *Brain and Language* 123. 174–182.

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

Almor, A., D. V. Smith, L. Bonilha, J. Fridriksson & C. Rorden. 2007. What is in a name? Spatial brain circuits are used to track discourse references. *NeuroReport* 18. 1215–1219.
Arbib, M. A. 1972. *The metaphorical brain: An introduction to cybernetics as artificial intelligence and brain theory*. New York: Wiley-Interscience.
Arbib, M. A. 2005. Beware the passionate robot. In J.-M. Fellous & M.A. Arbib (eds.), *Who needs emotions? The brain meets the robot*, 333–383. New York: Oxford University Press.
Arbib, M. A. 2006. Aphasia, apraxia and the evolution of the language-ready brain. *Aphasiology* 20. 1–30.
Arbib, M. A. 2007. Autism – More than the mirror system. *Clinical Neuropsychiatry* 4. 208–222.
Arbib, M. A. 2008. From grasp to language: Embodied concepts and the challenge of abstraction. *Journal of Physiology-Paris* 102. 4–20.
Arbib, M. A. 2011. Niche construction and the evolution of language: Was territory scavenging the one key factor? *Interaction Studies* 12. 162–193.
Arbib, M. A. 2012. *How the brain got language: The Mirror System Hypothesis*. Oxford: Oxford University Press.
Arbib, M. A. In press. Neurolinguistics. In B. Heine & H. Narrog (eds.), The Oxford handbook of linguistic analysis. Oxford: Oxford University Press.
Arbib, M. A. & M. Bota. 2003. Language evolution: Neural homologies and neuroinformatics. *Neural Networks* 16. 1237–1260.
Arbib, M.A., V. Ghanesh & B. Gasser. 2013. Dyadic Brain Modeling, Ontogenetic Ritualization of Gesture in Apes, and the Contributions of Primate Mirror Neuron Systems. *Phil Trans Roy Soc B*. to appear.
Arbib, M. A. & J. M. Fellous. 2004. Emotions: From brain to robot. *Trends in Cognitive Science* 8. 554–561.
Arbib, M. A. & A. Iriki. 2013. Evolving the language- and music-ready brain. In M. A. Arbib (ed.), *Language, music, and the brain: A mysterious relationship, Strngmann Forum Reports*, vol. 10, 359–375. Cambridge, MA: MIT Press.
Arbib, M. A. & G. Rizzolatti. 1997. Neural expectations: A possible evolutionary path from manual skills to language. *Communication and Cognition* 29. 393–424.
Aziz-Zadeh, L., T. Sheng, S.-L. Liew & H. Damasio. 2012. Understanding otherness: The neural bases of action comprehension and pain empathy in a congenital amputee. *Cerebral Cortex* 22. 811–819.
Baddeley, A. 2007. *Working memory, thought, and action oxford*. Oxford: Oxford University Press.
Barrès, V. & J. Y. Lee, 2013. From visual scenes to language and back via Template Construction Grammar. *Neuroinformatics*, DOI 10.1007/s12021-013-9197-y.
Bellugi, U. & S. Fischer. 1972. A comparison of sign language and spoken language. *Cognition* 1. 173–200.
Bickerton, D. 2009. *Adams tongue. How humans made language, how language made humans*. New York: Hill & Wang.
Buccino, G., F. Lui, N. Canessa, I. Patteri, G. Lagravinese, F. Benuzzi, C. A. Porro & G. Rizzolatti. 2004. Neural circuits involved in the recognition of actions performed by nonconspecifics: An FMRI study. *Journal of Cognitive Neuroscience* 16. 114–126.
Chen, Y., S. Fu, S. D. Iversen, S. M. Smith & P. M. Matthews. 2002. Testing for dual brain processing routes in reading: A direct contrast of chinese character and pinyin reading using fMRI. *Journal of Cognitive Neuroscience* 14. 1088–1098.
Cheney, D. L. & R. M. Seyfarth. 2005. Constraints and preadaptations in the earliest stages of language evolution. *The Linguistic Review* 22. 135–159.

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

Cheney, D. L. & R. M. Seyfarth. 2007. Baboon metaphysics: The evolution of a social mind. Chicago: University Of Chicago Press.
Chersi, F., P. F. Ferrari & L. Fogassi. 2011. Neuronal chains for actions in the parietal lobe: A computational model. *PLoS One* 6.e27652.
Cossu, G., S. Boria, C. Copioli, R. Bracceschi, V. Giuberti, E. Santelli & V. Gallese. 2012. Motor representation of actions in children with autism. *PLoS One* 7.e44779.
Dapretto, M., M. S. Davies, J. H. Pfeifer, A. A. Scott, M. Sigman, S. Y. Bookheimer & M. Iacoboni. 2006. Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience* 9. 28–30.
de Villiers, J. G. & P. A. de Villiers. 2009. Complements enable representation of the contents of false beliefs: The evolution of a theory of theory of mind. In S. H. Foster-Cohen (ed.), *Language acquisition*, 169–195. New York: Palgrave Macmillan.
Deacon, T. W. 1997. *The symbolic species: The co-evolution of language and the brain*. New York: WW Norton.
Deacon, T. W. 1998. Language evolution and neuromechanisms. In W. Bechtel & G. Graham (eds.), *A Companion to Cognitive Science,* 21225. Malden, MA: Blackwell.
Deacon, T. W. 2006. The aesthetic faculty. In M. Turner (ed.), *The artful mind: Cognitive science and the riddle of human creativity*, 2153. Oxford: Oxford University Press.
Dominey, P. F., M. Hoen & T. Inui. 2006. A neurolinguistic model of grammatical construction processing. *Journal of Cognitive Neuroscience* 18. 2088–2107.
Donald, M. 1991. Ori*gins of the modern mind: Three stages in the evolution of culture and cognition*. Cambridge, MA: Harvard University Press.
Dubreuil, B. 2010. Paleolithic public goods games: Why human culture and cooperation did not evolve in one step. *Biology & Philosophy* 25. 53–73.
Dunbar, R. 1993. Co-evolution of neocortex size, group size and language in humans. *Behavioral and Brain Sciences* 16. 681–735.
Edwards, D. H. & N. Spitzer. 2006. Social dominance and serotonin receptor genes in crayfish. *Current Topics in Developmental Biology* 74. 177–199.
Emmorey, K., J. Xu, P. Gannon, S. Goldin-Meadow & A. Braun. 2010. CNS activation and regional connectivity during pantomime observation: No engagement of the mirror neuron system for deaf signers. *Neuroimage* 49. 994–1005.
Fay, N., M.A. Arbib & S. Garrod. 2013. How to Bootstrap a Human Communication System. *Cognitive Science*, DOI: 10.1111/cogs.12048.
Fay, N. & S. Lim. 2012. From hand to mouth: An experimental simulation of language origin. In A. D. M. Smith, M. Schouwstra, B. de Boer & K. Smith (eds.), *The evolution of language*, 401–402. Singapore: World Scientific.
Feakins, S. J., P. B. de Menocal & T. I. Eglinton. 2013. Biomarker records of late Neogene changes in northeast African vegetation. *Geology* 33. 977–980.
Ferrari, P. F., V. Gallese, G. Rizzolatti & L. Fogassi. 2003. Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience* 17. 1703–1714.
Fogassi, L. & P. F. Ferrari. 2012. Cortical motor organization, mirror neurons, and embodied language: An evolutionary perspective. *Biolinguistics* 6. 308–337.
Gallese, V., M. Rochat, G. Cossu & C. Sinigaglia. 2009. Motor cognition and its role in the phylogeny and ontogeny of action understanding. *Developmental Psychology* 45. 103.
Gasser, B., E. Cartmill & M. A. Arbib. In Press. Ontogenetic ritualization of primate gesture as a case study in dyadic brain modeling. *Neuroinformatics*.

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

Hecht, E. E., D. A. Gutman, T. M. Preuss, M. M. Sanchez, L. A. Parr & J. K. Rilling. 2012. Process versus product in social learning: Comparative diffusion tensor imaging of neural systems for action executionobservation matching in macaques, chimpanzees, and humans. *Cerebral Cortex* 23(5). 1014–1024.

Hickok, G. & D. Poeppel. 2004. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 92. 67–99.

Hwang, S.-O.K. 2011. *Windows into sensory integration and rates in language processing: insights from signed and spoken languages*. Ph.D. Dissertation, Department of Linguistics, University of Maryland.

Iriki, A. & M. Taoka. 2012. Triadic (ecological, neural, cognitive) niche construction: a scenario of human brain evolution extrapolating tool use and language from the control of reaching actions. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367. 10–23.

Iverson, J. M., O. Capirci & M. C. Caselli. 1994. From communication to language in two modalities. *Cognitive Development* 9. 23–43.

Jeannerod, M. 2005. How do we decipher others minds? In J.-M. Fellous & M. A. Arbib (eds.), *Who needs emotions: The brain meets the robot*, 147–169. Oxford: Oxford University Press.

Knapp, H. P. & D. P. Corina. 2010. A human mirror neuron system for language: Perspectives from signed languages of the deaf. *Brain and Language* 112. 36–43.

Kuhl, P. K. & J. D. Miller. 1975. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science* 190. 69–72.

Levinson, S. C. 2013. Cross-cultural universals and communication structures. In M. A. Arbib (ed.), *Language, music, and the brain: A mysterious relationship. Strngmann Forum Reports*, vol. 10, 67–80. Cambridge, MA: MIT Press.

Liebal, K., M. Carpenter & M. Tomasello. 2010. Infants' use of shared experience in declarative pointing. *Infancy* 15. 545–556.

Liberman, A. M. & I. G. Mattingly. 1989. A specialization for speech perception. *Science* 243. 489–494.

Meltzoff, A. N. 2005. Imitation and other minds: The like me hypothesis. In S. Hurley & N. Chater (eds.), *Perspectives on imitation*, 55–78. Cambridge, MA: The MIT Press.

Menenti, L., S. C. Garrod & M. J. Pickering. 2012. Towards a neural basis of interactive alignment in conversation. *Frontiers in Human Neuroscience* 6.0.3389/fnhum.2012.00185.

Moulin-Frier, C. & M. A. Arbib. 2013. Recognizing speech in a novel accent: The motor theory of speech perception reframed, DOI 10.1007/s00422-013-0557-3.

Oztop, E., D. Wolpert & M. Kawato. 2005. Mental state inference using visual control parameters. *Cognitive Brain Research* 22. 129–151.

Padden, C. A. 2000. Simultaneous Interpreting across Modalities. *Interpreting* 5. 169–185.

Penn, D. C. & D. J. Povinelli. 2007. On the lack of evidence that non-human animals possess anything remotely resembling a theory of mind. Philosophical *Transactions of the Royal Society B: Biological Sciences* 362. 731–744.

Pickering, M. J. & S. Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*. 36. 329–347.

Prather, J. F., S. Peters, S. Nowicki & R. Mooney. 2008. Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* 451. 305–310.

Russon, A. & K. Andrews. 2010. Orangutan pantomime: elaborating the message. *Biology Letters* 7. 627–630.

Seyfarth, R. M., D. L. Cheney & T. J. Bergman. 2005. Primate social cognition and the origins of language. *Trends in Cognitive Sciences* 9. 264–266.

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press

Skipper, J. I., S. Goldin-Meadow, H. C. Nusbaum & S. L. Small. 2007. Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language* 101. 260–277.
Stanford, C. B. 1999. *The hunting apes: Meat eating and the origins of human behavior*. Princeton: Princeton University Press.
Steels, L. 2003. Evolving grounded communication for robots. *Trends in Cognitive Sciences* 7. 308–312.
Stout, D. 2011. Stone toolmaking and the evolution of human culture and cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366. 1050–1059.
Syal, S. & B. L. Finlay. 2011. Thinking outside the cortex: Social motivation in the evolution and development of language. *Developmental Science* 14. 417–430.
Tomasello, M. 1999. The human adaptation for culture. *Annual Reviews of Anthropology* 28. 509–529.
Vigliocco, G., D. P. Vinson, J. Druks, H. Barber & S. F. Cappa. 2011. Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience Biobehavioral Reviews* 35. 407–426.
Waddington, C. H. 1953. The "Baldwin Effect", "Genetic Assimilation" and "Homeostasis." *Evolution* 7. 386–387.
Waddington, C. H. 1957. *The strategy of the genes*. London: Allen & Unwin.
Whitehead, C. 2010. The culture ready brain. *Social Cognitive and Affective Neuroscience* 5. 168–179.
Williams, J. H. G. 2008. Self-other relations in social development and autism: Multiple roles for mirror neurons and other brain bases. *Autism Research* 1. 73–90.
Williams, J. H. G., A. Whiten, T. Suddendorf & D. I. Perrett. 2001. Imitation, mirror neurons and autism. *Neuroscience Biobehavioral Reviews* 25. 287–295.
Zukow-Goldring, P. 2012. Assisted imitation: First steps in the seed model of language development. *Language Sciences* 34. 569–582.

https://doi.org/10.1515/langcog-2013-0020 Published online by Cambridge University Press