Behavioral and Brain Sciences

Author's "Electronic Response" to Supplemental Commentaries

The mirror system hypothesis stands but the framework is much enriched, Part 2

Michael A. Arbib

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

This "electronic response" (ER) is the continuation of my Response (R) to those commentaries on my target article (TA) "From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics" published with the target article in *Behavioral and Brain Sciences* 28:2, 105-167. Some aspects of these commentaries have been explicitly addressed in the original Response. Where appropriate, I have explicitly labeled ER sections as continuations of R sections. This "electronic response" (ER) is directed at the 15 Supplemental Commentaries which have been posted at the following URL: http://www.bbsonline.org/Preprints/Arbib-05012002/Supplemental/.

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

ER1. More on Mirror Neurons

ER1.1. Are Mirror Neurons Special?

Pulvermüller uses the term "echo neurons" for neurons linking vocalization sounds and syllables to their respective motor programs. He speculates that these may have evolved together with, not subsequent to, mirror neurons for manual activities, arguing that both "exploit the same neuroscientific principles". He sees mirror neurons for grasping as important because they reveal functional principles immanent to the nervous system, but argues that they do not explain language. I agree with part of this. In R2.1, I hypothesized that the mirror system for grasping evolved in two stages: (a) to provide feedback for dexterous manual control, then (b) to underwrite the ability to act with other brain regions to make information available for interacting with others. Stage (a) implicitly invokes the general principle that the evolution of a motor system implies evolution of a supporting feedback system. Clearly, having the "feedback property" does not imply that one has evolved from, or is the evolutionary precursor to, the mirror system for grasping. However, one must have the "evolutionary pressure" to develop the skill that makes this feedback

adaptive. Since many nonhuman primates fare well with closed call systems, the evolutionary pressure to develop speech directly from such calls is unclear. I thus advanced (TA6.1) the view that the path to protospeech is built upon the scaffolding afforded by the evolution of the mirror neurons for grasping to provide a mirror system for protosign.

Pulvermüller holds that when words are frequently used together with objects, neurons in the ventral visual processing stream along with neurons in inferior frontal cortex responding to the visual features of the object may become active with perisylvian word-related networks, thus binding word and object representations to each other. This linkage is akin to the linkage of word and concept diagrammed in TA Figure 4; it is orthogonal to the claim that "the mirror for words" evolved on the basis of the scaffolding provided by the mirror system for grasping and, in turn, protosign.

ER1.2. The Plasticity of Mirror Neurons

Piattelli-Palmarini & Bever speak of "The kind of mysterious interiorization processes once invoked by Piaget" and remind us of the "failure" of Piaget's project of deriving linguistic structures from motor schemas. I reject the characterization of Piaget's project as a failure. He made great contributions to our understanding of the child's "construction of reality" and I have learned much from him – accepting certain ideas, learning from but correcting others (see Arbib, Conklin and Hill, 1987). At UMass, in the early 1980s, I conducted a graduate seminar called "Beyond Chomsky and Piaget" based on discussion of Piattelli-Palmarini (1980). I was struck by how well the two scientists represented their creeds in their respective chapters. Chomsky, the innatist, had all the answers – Piaget was clearly wrong and not rigorous, and that was that. Piaget, whose theory included learning by accommodation, on the other hand, was happy to admit that Chomsky had some valuable points to make, and did not have the discourtesy to note the impropriety of Chomsky, his linguistics founded on a competence theory of syntax explicitly divorced from language use, being so dismissive of insights gained from years of thoughtful observation of children. I do not believe that a rigorous theory of very little is inherently superior to a preliminary theory of very much.

Piattelli-Palmarini & Bever assert that is not known at present whether mirror neurons are as a consequence of a genetic program, or whether some canonical neurons can "turn into" mirror neurons by means of learning and experience. This seems mistaken. Mirror neurons and canonical neurons are anatomically segregated and have distinct patterns of parietal input. My claim is not that canonical neurons become mirror neurons but that mirror neurons attain their functional specialization through experience. This is attested to by the fact that a monkey can acquire mirror neurons for tearing paper and for tool use. Piattelli-Palmarini & Bever ask "If nothing at all is 'prewired' how can a mirror neuron do anything at all with the data out there, even determine that it constitutes experience for X? … Suppose, as Arbib suggests, that neurons become mirror neurons solely on the basis of experience. Then one should expect that a mirror neuron may be produced that is sensitive to the macaque's own action of grasping and, say, someone else's action of yawning." This has nothing to do with what is written in the TA! I say explicitly that the input to the MNS model is the *hand state*, which relates observation of the hand relative to the affordances of the object. This focus on hand-object relationships is "innate" – but the nature of the grasps that will be learned depends on the specificities of experience.

Piattelli-Palmarini & Bever rightly stress that learning models need appropriate criteria of equivalence, relevance and similarity. Indeed, there is a body of results in the neural networks literature (Arbib, 2003) on the way in which the generalization properties

of a network will depend both on properties of the network and the statistical ensemble of its inputs. Piattelli-Palmarini & Bever assert that similarity must be "inside the scope of the thought (must be a component of the perception itself)". I disagree. Once one considers the praxis of an organism, one has external criteria of efficacy. If, for example, I observe someone else acting to achieve a goal I find desirable, I may try to imitate the action to achieve that goal myself. If my action fails dismally, I have learned too little. If my action almost succeeds, then by trial and error I may tune that action to eventually gain the ability to successfully attain the goal. I made the move from "pure perception" to "action-oriented perception" more than 30 years ago (Arbib, 1972).

ER2. Internal Models

ER2.1. Coordinated Control Programs Must Include Goals

Makino et al. observe that success in complex imitation requires the ability to recognize the goal of an action as the basis for mastering the action which achieves that goal. Indeed, the MNS model (TA Figure 3) recognizes an action in terms of the goal of successful grasping of a specific affordance. Complex imitation further requires recognition of the way in which different actions fit together to achieve subgoals of the overall goal. This is akin to what Wohlschläger, Gattis, & Bekkering (2003) call goal-directed imitation: the imitator decomposes the observed movement into hierarchically ordered aspects with the highest aspect serving as the imitator's main goal. However, De Renzi (e.g., 1989) finds that some apraxics have difficulty both in classifying gestures and in performing familiar gestures on command yet may be able to copy a movement without "getting the meaning of the action" of which it is part. Let me call this *low-level imitation* to distinguish it from imitation based on recognition and "replay" of a goal-directed action. Failure to achieve a goal may cause us to attend to details of the movement that may constrain the choice of, or amend, a motor schema – "Oh, he didn't just grasp the handle; he gave it a twist as he grasped it." Indeed, human motor skill training may rest as much on such low-level imitation of an effector's trajectory as on attention to subgoals.

Such an observation is relevant to analyzing **Piattelli-Palmarini & Bever's** assertion that "single-trial emergence of novelty by activation is a cornerstone of nativist-selectivist models" but that "they remain utterly mysterious in all externalist gradualist models like Arbib's." I confess to puzzlement, perhaps because I have developed a specific theory rather than sought to sail under broad banners like "externalist gradualism", whatever that may be. The idea of complex imitation is that new skills are mastered by a combination of recognizing the structure of an action in terms of known actions (whether in one trial or several) and much practice ("gradualism?") to hone that skill. Pantomime is an example of a "one-off" strategy for conveying meaning – there is nothing gradualist about it.

Fagg and Arbib (1992) modeled the surprising speed with which monkeys could learn the association between a visual pattern and a motor response (Mitz, Godshalk and Wise, 1991). This led me to speak of Stage I versus Stage II learning. In Stage I, which may take months of shaping, the monkey learns the general task "associate a new visual pattern on this screen with the correct pull on the lever in front of you and you will get a reward". In Stage II, he takes only 7 or so trials to stabilize the correct response to each novel visual pattern. The issue here is how one can go (evolution) from an ancestral brain for which Stage I learning can never yield the ability to learn language to a human brain in which perhaps 2.5 years of learning is required for Stage I to make possible the "word explosion" which distinguishes the human infant from any nonhuman primate.

ER2.2. Internal Models and the Cerebellum

Ito argues that when the motor cortex moves a hand repeatedly responding to a plan and referring to sensory feedback of the actual hand movement, an internal model simulating the dynamics of the hand will be formed in the cerebellum. Ito views the internal model mechanism of the cerebellum as separate from the mirror neuron mechanism, with the two cooperating. I agree with the view implicit in Ito's Figure 1 that the cerebellum functions as a side path to control by motor cortex, but would argue (since Holmes, 1939, showed that a plan can be created without the cerebellum but that its execution is then terribly impaired.) that it provides additional tuning to a model encoded in cerebral cortex, rather than being the only model for an action (Arbib, Érdi & Szentágothai, 1998). "Sensory System + F5" in Ito's figure provides a highly compacted view of the brain regions detailed in the FARS and MNS models (TA Figures 2 and 3). I have suggested, though, that the macaque mirror system does not support imitation to any great extent, and thus FARS+MNS requires extension to account for how observing an action may trigger its repetition, and how complex imitation (goal-directed imitation) can install a model of a compound movement which is then subject to development of a corresponding cerebellar model to better coordinate the subactions involved. In the case of "one-off plans", presumably cerebellar tuning can apply to familiar components and transitions but cannot compensate for hesitations when unfamiliar choices or transitions must be made.

ER3. Etcetera (Continuing R1.2)

ER3.1. Lateralization

Nagy reports that human neonates in their first 3-96 hours of life showed a left-sided bias during imitation but not during the baseline period. **Nagy** thus suggests that human neonates may use a specific lateralized brain system for imitation. However, this is the opposite to the laterality of most infants for dexterity and language and also seems contrary to the fact that young children spontaneously imitate adults in a mirror-like fashion (left hand imitates right hand, etc.) whereas older children tend to transpose left and right (Wohlschläger et al. 2003). Perhaps right-handed parents tend to extend their right limb and thus infants more often respond with their left limb – and this bias persists even on those occasions when the imitatee uses the left hand. Since just waving an arm is unskilled, the "normal laterality" may only come into play when the child is much older and has to start being dexterous.

ER3.2. Brain Evolution

Ito stresses that a change in genetic code responsible for formation of neuronal circuitry underlies each stage of the evolution of brain function: addition of a new brain area elaborates the control system structure of neuronal circuits, and consequently its operation; while introduction of novel cellular or molecular processes may modify the architecture, and consequently the operation, of a local neuronal circuit. I have noted the cerebellum's role as crucial to the graceful adaptation and coordination of movements, but with the plan of movement elaborated elsewhere, while Ito has noted that lateral cerebellum can contribute to even highly cognitive aspects of neural function. I speculate that mammalian

cerebral cortex and cerebellum are coupled in a genetically controlled system in such a way that they can co-evolve.

ER4. Imitation in Developmental Perspective (Continuing R2.2)

Nagy asks why babies start to imitate and what is the motivation for their first imitations. Social reciprocity in neonatal imitation may be a necessary precursor for complex imitation. **Nagy** further suggests an innate basis for conversation that precedes its pragmatic function (see R3.1). The caregiver acts as a "mirror" (not in the mirror neuron sense) to help the child place his own actions in social context – the genetics of both infant and caregivers cooperate to provide the conditions for "conversation" as part of complex social learning.

Zukow-Goldring glosses TA as suggesting that communicative gestures emerged eventually from a shared understanding that actions one makes oneself are indeed similar to those made by conspecifics, establishing that "I am like the other". That is, this common understanding of action sequences may provide another "missing link" to language – where action *sequences* rather than single actions may be crucial. Zukow-Goldring does not deny that children can learn certain things for themselves by trial-and-error. However, by directing the child's attention to its own effectivities in relation to affordances in the environment, the caregiver greatly narrows the search space for learning, and consequently enhances the speed and extent of that learning.

ER4.1. Babbling

Pulvermüller stresses that humans go through a stage of vocal and manual babbling during which arbitrary motor patterns are practiced extensively and that this stage "which must have a specifically human genetic basis" sets up a great variety of action-perception networks. This seems more relevant to building the phonological stock than to the child's vocabulary explosion. Perhaps (to return to my earlier terminology) babbling is *part of* Stage I for the Stage II vocabulary explosion, but only for the motor aspects of words, not for mastery of their meanings.

Horwitz et al. observe that the MNS model may require considerable modification when the focus shifts to the auditory modality because, for instance, there is no treatment of babbling. This is true as it stands, but MSH does seek to show how mechanisms for manual control might have evolved to support vocal control. The ILGM model (see discussion attendant to TA Figure 3) might be seen as a theory of how "manual babbling" leads to an effective set of grasps, and Arbib & Rizzolatti (1997) discussed the relevance of inverse and forward models, building on the insights of Jordan and Rumelhart (1992) in thinking about the infant's vocal babbling. Fatima Husain (p.c.) adds that babbling is important in that it can be used to gain insights into speech production (Locke 1993) and into the evolutionary origins of language (Lieberman 2000).

ER5. Theory of Mind (Continuing R3.3)

Makino et al. argue that parity – what counts for the speaker must count for the listener – "plays a much more crucial role in communication than can be inferred from [this definition]", namely the role of "understanding others". This observation seems more an example of what could not be covered in the space of the TA than a critique of MSH – see the related discussion in R3.3. Successful communication involves prediction of each

other's action, which requires estimation of each other's mental state. In relation to this, Fogassi et al. (2005) showed that activity of neurons (both motor and mirror) in the inferior parietal lobule (IPL) may correlate not so much with the current subgoal as with the monkey's "intention" – the overall behavior of which the current action is just a part. When monkeys grasped a piece of food prior to eating it versus prior to placing it in a container, activity of IPL motor and "mirror" neurons coding a specific act was very different depending on the overall action of which it was part. The cue that determines the overall action is the presence or absence of the container – if the container is present the food is to be placed in it; if not, the food is to be placed in the mouth.

Makino et al. offer to improve the "intended communication hypothesis" by assuming parity. If the utterer is able to estimate the recipient's mental state, the utterer can recognize a pantomime's effect on the recipient in terms of the change of the recipient's estimated mental state. However, I think they may have this the wrong way round. I am trying to understand the transition from the monkey, for which pantomime cannot convey meaning, to a human, for which it can. Here, we cannot *assume* parity, we must *explain* how communication parity emerged from that for praxis. Makino et al. further suggest that complex imitation can be optional for intended communication. They are right for simple forms of communication but complex imitation is needed for the capacity to acquire *language*, going from recognition of isolated gestures to generation and recognition of complexes of gestures – with syntax becoming required as more and more complex structures becomes codable by linear strings of symbols.

Makino & Aihara (2003) analyzed John's need to know the dynamics of Mary's mental state and its relation to her action and environments within the framework of "dynamics estimation".**Makino et al.** show how the difficulty of achieving *mutual* estimation of each other's mental state within its high-dimensional mental state space is solved by exploiting similarities between peers. This is reminiscent of the simulation theory of mind-reading (e.g. Gallese & Goldman, 1998). Just as the MNS model correlates externally observable features (the visual trajectory of hand relative to object affordance) with unobservable internal features (the encoding of an action by F5 canonical neurons), so can John observe himself and create an internal model (simulation model) for his own behavior. However, this leaves unresolved the "binding problem", which is akin to the creation of schema instances. We may have generic knowledge of chairs embedded in a single *chair schema*, but when looking at a room, we may need to maintain several *instances* of the chair schema, each furnished with the particulars of the location and characteristics of the corresponding observed chair. Similarly, we must link an appropriate instance of the model to each agent and keep separate the actual and predicted mental states for oneself and the others with whom one is interacting.

ER6. Biological Neural Networks (Continuing R4.1)

ER6.1. Neural circuits for syntax

Pulvermüller has developed a model of language processing in which words are represented by distributed cell assemblies whose cortical topographies reflect aspects of word meaning and opts for specifying the putative neural circuits of syntax and recursion first and then thinking about the possible evolution or emergence of these mechanisms. That's a reasonable strategy, but I have two reservations:

(a) By focusing on syntax, one will develop a model that ignores links between action and language that may (as MSH asserts) be crucial to understanding evolution.

(b) Turing to theory of computation, we know that if we bound the nesting in a grammar, we can implement it in a finite automaton, and we have known since 1943 (or at least 1956) that we can implement an automaton as a formal neural network. **Pulvermüller** enriches this by showing how to use associative learning rather than ad hoc construction to build these networks. Nonetheless, his very interesting models make no contact with data on real neural circuitry of the primate brain, and his book (Pulvermüller, 2003), treats very little syntax. Relative clauses are mentioned in passing and questions seem not to be studied at all. (Some of these limitations are being removed in current work.)

Given this, I do not see how such modeling helps us "think about the possible evolution or emergence of these mechanisms".

ER6.2. Functional Neurophysiology of Argument Structure

Where the TA simply asserts that "It might have taken many, many millennia for people to discover syntax and semantics in the sense of gaining immense expressive power by 'going recursive' with a relatively limited set of strategies for compounding and marking utterances.", **Dominey** provides a welcome computational approach to providing a functional-neurophysiological characterization of how this could happen. Dominey focuses on sequence processing, whereas the essence of sentence structure is its buildup from hierarchical constituents. I am not yet convinced by Dominey's claim (p.c.) that the sequence problem must be solved before one can start worrying about more complex structures. In response to my query "What are the function symbols in sensorimotor function", Dominey (p.c.) speculates that these will include saccades, that punctuate the construction of multidimensional space, and features common to two effector motor sequences used to accomplish the same goal. This seems consistent with building up hierarchical behavior rather than a prior drive related solely to sequence learning.

ER7. Processing Manual and Linguistic-Prosodic Gestures (Continuing R5.3)

Rogalewski et al. provide data for a joint neural system for the processing of manual and linguistic-prosodic gestures. Together, these data indicate that only vocalizations with a linguistic component – and not nonspeech vocalizations – are functionally linked with the manual gesture system. Rogalewski et al. see this as in agreement with my view that language evolution proceeded from species-specific vocalizations to manual gestures and thence to protospeech and full-blown language. I welcome this support but would like to see further argument linking the data on current human brains to the supposed evolutionary progression.

ER8. For and Against Holophrasis (Continuing R6.1)

Rules of combination are synthetic, putting atomic terms together to make larger expressions. Syntax is thus essentially synthetic by definition. The issue is whether this

synthetic state of affairs was preceded – as the TA argues – by a period when the reverse "analytic" process was dominant. For **Hurford**, the synthetic account is the simpler account. By what measure? One might say that "phonetic writing systems are simpler than ideographic writing systems" but, again, by what measure? Starting from pictures, it is false. Starting from sound patterns, it is true. Similarly, once you have discovered the combinatorial power of using syntax to combine words, then words are simpler. But if you have not done so, then labeling significant events or constructs seems the simpler strategy – which is why I argue that the analytic story deserves greater attention rather than out-of-hand rejection.

Hurford and **Piatteli-Pakmarini & Bever** note that Wray (2000) notes that holistic expressions are still very common in modern languages and sees them as a carry-over from holistic pre-modern communication. They counter that current holistic expressions arose through a process of historical fossilization of previously productive constructions. However, Wray is not asserting that the holophrases *of protolanguage* were compounds of words – quite the contrary – but rather shows that holistic expressions are as much a part of modern language as analytic expressions, and thus that we should not dismiss their importance for protolanguage. We are trying to imagine hominids without syntax and understand how they got syntax. I don't deny that some protowords might be akin to modern words. What I do deny is that every utterance, forever in the prehistory of mankind, was indeed an aggregate of words as we know them. I think part of the problem is that people have trouble thinking through the fact of Deep Time. There are many "self-evident truths" that were not always self-evident, and it is a feat of the imagination to think back to a possible past in which the culture had not yet made them self-evident.

Hurford (2003) argues that a simple whole meaning, of the kind destined to find eventual linguistic expression, has two parts, a predicate and an argument, formally PREDICATE(x) and that these two separate semantic elements can be identified with separate neural pathways. These claims are interesting but controversial – I do not have space to analyze them here. But they predispose **Hurford** to "suspect that the penchant for [holophrasis] comes from a tacit acceptance of the schoolbook idea of defining sentences as expressing whole meanings. The idea of a 'whole meaning' is never satisfactorily defined independently of grammatical subject-predicate structure, which makes its use in the definition of sentence hopelessly circular." I disagree. (i) I may use differential equations to describe the orbit of a planet without implying that planets know calculus. Similarly, there is nothing circular in using a formalism to describe the behavior of creatures who are ignorant of that formalism. (ii) I explicitly go far beyond just two parts. Early humans might have had a single protoword that means "The great chief has killed a large meat animal and we shall feast tonight. Yum, yum." with no words to express the separate parts. (iii) Hurford confuses the assertion that a statement has two parts in a current language with the idea that its two parts must be explicit in all representations. If I ask you to open a door, you will reach for the handle, turn it, and pull open the door in one integrated movement. The handle is implicit in the performance but not explicit. In the same way, certain signs of ASL combine object and action in one overall gesture. The fact that logicians can describe something with subject-predicate structure does not mean that the structure is explicit in the describee (a point made in my critique of Hurford's 2003 BBS article). Indeed, one just has to look at translations from one modern language to another to see that a word in one language may only be translatable by a phrase in another – one tribe's word is another's protoword.

Piattelli-Palmarini & Bever cite instances of the vast literature on the compositionality of the semantics of natural languages, and note that compositionality is intimately tied to

productivity and systematicity. They then assert that "Short of being another miracle to be added to Arbib's list, there is no conceivable continuity, conceptual or factual, between a communication system based on monolithic conventional signs and a fully compositional language." But this would seem to assert – contra Bickerton as much as contra me – that there was no such thing as protolanguage, since neither version invokes compositionality in its full sense. They then assert "Noncompositional languages are un-learnable, each symbol-meaning pair has to be memorized (the need of an explicit convention for each expression is a result of that)." By the same logic, the lexicon is unlearnable. We must now add the innate Universal Lexicon to Piattelli-Palmarini & Bever's set of counter-miracles!

Piattelli-Palmarini & Bever assert that fractionation of monolithic expressions presupposes an access to internal principles of phonology, morphology and syntax and cannot create these principles from scratch. However, their view treats phonology, morphology and syntax as indivisible entities that cannot exist without all their principles in place. I reject this, arguing that each emerged gradually by a process of bricolage. We have an archeological and written record that lets us trace the cumulative inventions that led to agriculture, writing, and living in cities over the last 10,000 years or so. Just because we lack an appropriate archeological record does not preclude that languages, too, rested on such cumulative inventions. It is because of this I took pains to make clear in TA that one's account of the evolution of the human brain might be seen as having two very different results: "the language-ready brain" versus "the brain that 'has' language." I find no compelling argument in Piattelli-Palmarini & Bever's commentary against my hypothesis that it is the former that evolved. Indeed, they seem to reject the possibility of any genuinely evolutionary account and write like Creationists – dogs beget dogs, human beget humans, we are so different from other species we must have been created as we are with every aspect of language in place.

ER9. Bringing in Semantics (Continuing R6.2)

ER9.1. Concepts and Actions

Hurford usefully discusses concepts and their representation in the brain. He notes my TA comment that only rarely (as in the case of certain basic actions such as *grasp* or *run* or certain expressions of emotion) will perceptual and motor schemas be integrated into a "mirror schema". He responds that many other verbs correspond to actions within the human repertoire. I do not deny that many verbs describe human actions for which mirror properties are highly likely but note that, for example, they would constitute a small fraction of the verbs in this article.

How can concepts be acquired? **Choe** offers a simple thought experiment to make the points that (1) voluntary action can provide meaning to one's internal perceptual state, and (2) maintained invariance in the internal perceptual state can serve as a criterion for learning the appropriate action sequence. Interestingly (cf. Arbib, 1997) the idea has quite a pedigree. Room (1967) showed how the experience that a farmer gains in learning how to lay out a field and plow regular furrows might eventually lead to systematic observations that could be encapsulated in the axioms of Euclidean Geometry. Nicod (1970) imagined how different creatures with different sensorimotor systems might induce quite different geometrical structures for their worlds – contrast the sensorimotor repertoire of our farmer with that of a creature confined to running up and down a piano keyboard with the sound of the notes thus struck providing its only sensory input. What we add here is that there is no limit of "one space, one creature"; rather each creature is capable of a multiplicity of

tasks, and each may deploy a number of different spatial representations in adapting task execution to varying circumstances.

Choe sees the TA example of grasping an apple, biting an apple, and so on, as showing how much a concept is defined by a network of expectations associated with actions for which there are mirror schemas. However, having said this, we have to note that there are perceptual schemas rather removed from actions - such as recognizing that something is "red" or "sweet". There is something of recursion in the learning here – e.g., by recognizing what I need to do to place something called "sugar" in my mouth I can enrich my understanding of "sweet" which then becomes a defining characteristic of "sugar". Similarly, my recognition of "green" may tie in with pairing the concept of "apple" with the taste called "tart" and this in turn may enrich my concept of "apple". (Cf. the "schema encyclopedia" of Arbib, 1985).

ER9.2. Word Acquisition in Children

Werning highlights the *"complex first" paradox* – the fact that substance concepts are acquired ontogenetically earlier than apparently more primitive attribute concepts. However, this may be misleading. For the child, the redness of his toy truck is as indissoluble as the fact that it has wooden wheels – so the word truck implies the "redness" and the "wheels" when he has words for neither. "Milk" at the one word stage may mean "gimme milk". Further, "hot" (German "heiss"), is among the first 15 words and is typologically an adjective, although it might not be used as such by the child (Werning, p.c.). Anyway, Werning sees this as contrasting with widely held views (e.g. Barsalou, 1999) that prototypical (first-level) attribute concepts are semantically primitive and the prototypical substance concepts are semantically complex with, e.g., the substance concept [*mango*] made up of the vector of attribute concepts <*orange, oval, big, soft, sweet, edible,* ...>. I advocate a schema semantics which is quite different. We may recognize a mango purely by its visual cues which then create expectations such as "if I squeeze it, it will be soft" and then, if this expectation is met, "If I bite into it, it will be sweet." There is a network of perceptual, motor and linking schemas, connected to a network of language schemas. However, this in itself, does not resoilve Werning's paradox. I think the answer is that a "semantic primitive" need not first be made cognitively explicit and symbolized accordingly prior to the lexicalization of other concepts. The child may have a concept implicitly as part of the basis for recognizing an object or action in its world, yet not have direct access to it. Just as (TA3.1) we separate the "what" and "how" pathways and note that motor parameters/affordances can be used (with great precision) by the "how" pathway but cannot be declared without the "what" pathway, so we must understand that only certain representations in the brain can be "put into words". Thus response to an orange-colored patch in early visual processing must not be confused with the neural representation of "orange" as a concept that can be put into words. As Werning notes, this echoes some of the disacussion of holophrasis – one may have access to a complex concept without being able to put its components and their relationships into words: "Given that semantically complex words are evolutionarily prior such an interface towards semantics … must have evolved at an early stage in the evolution of language and it must have had strong distributive capacities from the beginning."

ER9.3. Evolution of Semantic Abilities

Dessalles & Ghadakpour stress that predicative structures should be considered a distinctive feature of the human communication system (contrary to Hurford, 2003; cf. ER8). Predication, negation and argument structures are unavailable to animals. I agree, but perhaps see more hope for an account of how to bridge that gap. Dessalles & Ghadakpour note that we can use language to systematically express the negative version of a predicative structure, e.g. "Leo doesn't grasp the raisin", while claiming there is no perceptive meaning corresponding to the negation of a visual scene. However, the effort to apply certain recognition criteria and having them fail to meet threshold justifies negation. Bennett (1964) provides a stimulating essay on what would have to be added to the "language" of bees – such as negation and the ability to weigh evidence – to arrive at creatures who were truly rational.

Dessalles & Ghadakpour suggest that I grant apes the same semantic representations as modern sapiens, but this is a misunderstanding. Even for those schemas which are shared between apes and humans, apes are unable (in general) to express such schemas in words. In my current account, protolanguage can associate symbols with the assemblage but cannot associate symbols with the elements of a structural description. Even a human cannot "reach into the brain" to provide a full assemblage description, but instead learns to recognize and express an increasingly rich set of relationships in the scene. However, I agree that it is a failing of the current account that it does not talk of social bonding, affect, expression of dominance, etc., let alone abstract thought and predicates. I see this as a challenge for future research on "The Invention of Languages" (TA7).

Recall three steps in TA: (i) our ancestors master pantomime; (ii) their followers develop the ability to associate conventional holistic signs to frequently encountered scenes; (iii) some *sapiens* invent compositional languages. Noting that conventional signs are claimed to fulfill a disambiguation purpose, **Dessalles & Ghadakpour** ask why we didn't evolve more powerful analogue imitating abilities? My answer is two-fold: (a) Economy of gesture leads to ritualization, with ease of expression winning out over analogical detail. (b) Some distinctions can be imitated whereas others cannot: e.g., bird vs. flying. Since there is no natural imitation of an object like raisin, we need a conventional gesture to distinguish action from goal, or the tracing of a characteristic outline to delineate the object. Whether for pantomime or compound words, the sign does not serve to unambiguously specify something, but rather to create a relevant state of mind. "Houseboat" and "housecoat" remind one of their component entities in very different ways – the former is "a boat that serves as a house" which might suggest that the latter is a tent ("a coat that serves as a house")! In this case, no compositional semantics is provided – rather we have two cues to retrieve a high-frequency assemblage to which they are related.

ER9.4. A Social Account of Semantics

Externalists deny that there is any fact about mental states or brain tissue that constitutes the meaning of words, insisting that meaning and language use depend crucially on public criteria and rely deeply on social norms. *Internalism* is the doctrine that features of the world are defined by states internal to current biological cognizers and that these states

hold the key to explaining cognition. **Ruchsow** reads many passages in TA as being in favor of an externalistic approach to meaning but sees some sympathy for internalism in my review of brain imaging studies. "Arbib oscillates between a compelling evolutionary approach of social language learning and internalistic arguments stemming from neuroimaging techniques." However, I would claim that I do not oscillate, I integrate! I regard both Externalism and Internalism as incomplete and believe that the challenge is to provide a framework in which the partial truths of each can be integrated. This is the challenge that Mary Hesse and I tackled (Arbib & Hesse, 1986). Mary addressed the issue that the facts underdetermine theory, so that any scientific theory is in some sense the social product of a community of scientists – but with the theory very much constrained by observation. I, on the other hand, wanted to understand the brain processes that could, for example, allow a creature to recognize an apple and eat it. Our solution was to expand my theory of "schemas in the head" to provide a complementary account of ""social schemas" implicit in the overt behavior of members of a community. I see our challenge as neuroscientists to be to understand how social norms or conventions become embodied – going from social to individual schemas. Ruchsow (p.c.) agrees with this aim, suggesting that it will serve to counterbalance the predominance of internalism in contemporary neuroscience. He believes this could lead to a pluralism of descriptive (natural sciences) and normative (social sciences) vocabularies, but welcomes such pluralism "as reductionism is always is in danger of some sort of fundamentalism".

ER10. Concerning Innate Universal Grammar (Continuing R6.3)

ER10.1. In Defense of Construction Grammar

The TA endorses the view that over the course of many generations of historical language transmission and change, grammatical systems gradually emerged and became increasingly complex. **Kemmerer** , in a way that I find convincing, shows that this position is supported by work of Croft (e.g., 2001), among others, which develops *construction grammars* to identify the unique constructions of each language, thereby respecting the diversity of grammatical systems. However, this approach is hospitable to cross-language comparison and can encompass a universal theory of grammatical categories grounded in semantic and pragmatic prototypes rather than isolated syntactic parameters. For example, cross-linguistically, prototypical nouns specify objects and have referential functions, prototypical verbs specify actions and have predicative functions, and prototypical adjectives specify properties and have modifying functions.

Kemmerer also charts the relevance of such ideas to language acquisition. He concludes that such considerations support the view that human languages contain a huge, open-ended spectrum of historically shaped, constructionally based, hierarchically organized, and distributionally learned grammatical categories.

ER10.2. In Defense of Innate Universal Grammar

Piattelli-Palmarini & Bever assert that "Michael Arbib nourishes a deep-seated and long-standing dissatisfaction (to put it mildly) with generative grammar and with almost everything it stands for." This assertion misrepresents my views, and is certainly not based on the TA. I first discussed grammars with Chomsky in 1961 and have since read many, many papers on generative grammar and learnt much therefrom. I welcome attempts by generative grammarians to describe syntactic structures, but reject the slide from a study of

syntactic competence to a pronouncement about what is innate to the human brain. Just as Kepler's work was essential to Newton's analysis of planetary dynamics, so can generative grammar describe patterns that challenge our understanding of language performance – but I deny that describing competence using parameters implies these parameters reside innately in the child's brain.

Piattelli-Palmarini & Bever are confident that "there is no such thing as a fraction of … a Noun Phrase or Verb Phrase." However, one may think of count nouns versus mass nouns, living nouns versus non-living nouns. the groups of nouns with their own classifier for counting as in Japanese, etc. It is a lack of imagination to doubt that early languages could exist without a unique category of Noun Phrase, rather than a set of categories that *we* would recognize as noun phrases. Further, the commentary by **Kemmerer** makes clear that different languages may have quite different sets of "Nouns" and "Verbs".

I have responded to the claim of Hauser, Chomsky and Fitch (2003) that recursion is unique to humans by noting that gorilla behavior exhibits recursive patterns on the basis that "if at first you don't succeed, try try again". **Piattelli-Palmarini & Bever** seem to concede this point but insist that I must supply "a language-specific hypothesis that describes exactly how recursion works in language". However, my point is many features of language do *not* require a language-specific explanation. Let me quote Arbib (2005) to see how viewing sentence production as goal-directed behavior may reveal some aspects of linguistic recursion as following naturally by applying the general "if at first you don't succeed" strategy in a new domain:

> [C]onsider a sentence like "Serve the handsome old man on the left.", spoken by a restaurant manager to a waiter. From a "conventional" linguistic viewpoint, we would apply syntactic rules to parse this specific string of words. But let us look at the sentence … as the result of the manager's attempt to achieve a *communicative goal*: to get the waiter to serve the intended customer. His *sentence planning strategy* repeats the "loop" <add adjective or prepositional phrase> until (he thinks) ambiguity is resolved:
>
> (1) Serve the old man.
> Still ambiguous?
>
> (2) Serve the old man on the left.
> Still ambiguous?
>
> (3) Serve the handsome old man on the left.
> Still ambiguous? Apparently not. So the manager "executes the plan" and says "Serve the handsome old man on the left." to the waiter.
>
> Here, a noun phrase NP may be expanded by adding a prepositional phrase PP after it [as in expanding (1) to (2)] or an adjective Adj before it [as in expanding (2) to (3)]. The suggestion is that syntactic rules of English which I *approximate* by NP $\rightarrow$ NP PP and NP $\rightarrow$ Adj NP are [here] abstracted from procedures which serve to reduce ambiguity in reaching a communicative goal … – and thus exemplifies ways in which reaching a pragmatic goal (identifying the right person, or more generally, object) may yield an unfolding of word structures in a way that may clarify the history of [recursion in] syntactic structures.

Piattelli-Palmarini & Bever dismiss my view that the development of syntax, semantics and lexical semantics was the result of multiple inventions by long-forgotten geniuses over the course of thousands of generations. But consider a parallel example. Digital computers did not exist before we were born; children of less than 5 can now use them. It took geniuses to invent computers and many innovations in hardware, software and networking to render them into a form that the child could use. It is highly presumptuous to look at an edifice (the range of human languages) that may have been 100,000 years in the making and assert that each and every feature was there "from the very start".

Piattelli-Palmarini & Bever insist that "Against practically every linguist's well-grounded persuasion, Arbib maintains that adjectives, verbs and nouns are not 'natural categories'" They are clearly wrong. TA footnote 25, for example, shows how different the notion of "adjective" may be in different languages. They then raise several interesting linguistic questions such as "Why not have verbs with 5 or more thematic roles?" I don't know, of course. But the way they set the stage for these specific questions seems to support my argument for cognitive grammar: languages may share certain communicative goals but differ in the syntactic strategies they use to achieve them. Unless one insists on the Universal Lexicon, one has conceded that language is inherently open to lexical innovation, and novel metaphors and idioms. Why then should one be resistant to syntactic innovation? **Piattelli-Palmarini & Bever** go on to say that the answer to their question is "exquisitely syntactic, in terms of the nature of available positions in the internal structure of the vP and the universally fixed number and configuration of these positions." But this is not an explanation, it is a description. Whether the explanation is rooted in biology or history is an open question.

Piattelli-Palmarini & Bever posit a conundrum: They postulate either that:

(a) the essence of language is "invented" culturally, as a result of human cognition. But this is equivalent to the claim that language is "innate" in a strong sense. Or that:

(b) the essence of language is "discovered", just as one might view the emergence of tools or of methods of food preparation – certain methods work because they interact felicitously with affordances of the physical world. Bever (1981) has claimed that the speed and characteristic universals with which children acquire language, shows that there must be innate mechanisms that direct the child to "discover" language quickly and in just the right way.

I do not follow their logic in (a). Consider "[If we] postulate that the essence of computer programming is 'invented' culturally, as a result of human cognition … this is equivalent to the claim that computer programming is 'innate' in a strong sense." The folly of this claim seems to invalidate assertion (a) about language.

As to (b), I note in TA that "it is not true that children master all the vocabulary or syntactic subtlety of a language by 5 or 7 years of age." The child does not "discover" language – or tools – in the same way that our distant ancestors did, creating new words and constructions to meet new communicative needs. The child learns the conventions of the community around it as it seeks to communicate. Such communication provides a pragmatic criterion quite different from explicit coaching in grammatical rules.

ER10.3. Description is Not Enough

Let me close by returning to **Piattelli-Palmarini & Bever**'s claim "Languages are all the same in important regards that transcend the universals of protolanguage." To the extent this claim is true, it is hardly counter to my general viewpoint. For example, the use of one class of words for actions and another for objects is nearly universal, and to the extent that it is it can be linked to the needs to communicate certain universal classes of cognitions rather than being rooted in grammar *per se*.

Baker (2001) provides perhaps the best and most accesible exposition of the use of a Universal Grammar based on parameters to capture many, often non-obvious, similarities between the syntax of different languages. He shows that languages can be surprisingly similar *if we strip away the vigorous differences in vocabulary and the way in which cultural differences affect the way sentences are formulated.* I read Baker's book both sympathetically and critically, and found myself impressed by how much Baker can describe about syntax in terms of his parameters, but also by how much of idioms and other constructions – essential to native command of a given language – lie outside his framework. Moreover, I see nothing in Baker's book to back the claim that an innate knowledge of parameters would ease language acquisition or help us understand historical patterns of language change (I hope to substantiate these conclusions elsewhere under the title, inspired by a conversation with Jean-Roger Vergnaud, "UG: Universal Grammar or Unidentified Gadgets?"). I believe that cognitive and construction grammar will, as claimed in the TA, provide a far better framework than Universal Grammar for describing language as a living medium for communication, and that it is as a living medium for communication that language can be seen to evolve from protolanguage in a way that can be illuminated by MSH.

References

Arbib, M. A., 1972,*The Metaphorical Brain: An Introduction to Cybernetics as Artificial Intelligence and Brain Theory*, Wiley-Interscience: New York.

Arbib, M. A., 1985, *In Search of the Person: Philosophical Explorations in Cognitive Science*, Amherst: University of Massachusetts Press

Arbib, M. A., 1997, Modeling visuomotor transformations, in *Handbook of Neuropsychology*, Volume 11, Section 16: *Action and Cognition*, (M. Jeannerod, ed.), Amsterdam: Elsevier, pp. 65-90.

Arbib, M.A., (Ed.), 2003, *The Handbook of Brain Theory and Neural Networks*, Second Edition, A Bradford Book/The MIT Press.

Arbib, M.A., 2005, A sentence is to speech as what is to action? *Cortex*, in press.

Arbib, M. A., and M. B. Hesse, 1986, *The Construction of Reality*, Cambridge: Cambridge University Press

Arbib, M. A. & Rizzolatti, G., 1997, Neural expectations: A possible evolutionary path from manual skill to language. *Communication and Cognition 2*9:393–424.

Arbib, M.A., Conklin, E.J., and Hill, J.C., 1987, *From Schema Theory to Language*, Oxford University Press.

Arbib, M.A., Érdi, P. and Szentágothai, J., 1998, *Neural Organization: Structure, Function, and Dynamics*, Cambridge, MA: The MIT Press.

Baker, M.C., 2001, The atoms of language: the mind's hidden rules of grammar. Oxford: Oxford University Press.

Barsalou, L.W., 1999, Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577-660.

Bennett, J., 1964, *Rationality: An Essay Towards an Analysis*, in Studies in Philosophical Psychology, R. F. Holland (ed), Routledge & Kegan Paul, The Humanities Press, New York.

Bever, T.G., 1981, Normal acquisition processes explain the critical period for language learning. In Diller, K. C. (Ed.), *Individual Differences and Universals in Language Learning Aptitude,* Rowley, Ma: Newbury House Publications, Inc., pp. 176-198.

Croft, W. (2001) Radical construction grammar. Oxford University Press.

DeRenzi, E. (1989). Apraxia. In F. Boller& J. Grafman (Eds.), *Handbook of neuropsychology.* Amsterdam: Elsevier. Vol. 2., pp. 245–263.

Fagg, A. H., and Arbib, M. A., 1992, A Model of Primate Visual-Motor Conditional Learning, *Adaptive Behavior,* 1:1-37.

Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S. , Chersi, F., and Rizzolatti, G., 2005, Parietal Lobe: from Action Organization to Intention Understanding, *Science*, 308: 662-667.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences, 2*, 493-501.

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

Holmes, G., 1939, The cerebellum of Man. *Brain* 62: 1-30.

Hurford, J.R., 2003, The neural basis of predicate-argument structure, *Behavioral and Brain Sciences* 26:261-283.

Jordan, M.I., and Rumelhart, D.E., 1992, Forward Models: Supervised Learning with a Distal Teacher, *Cognitive Science* 16:307-354.

Lieberman, P., 2000, *Human Language and our Reptilian Brain: The Subcortical Bases of Speech, Syntax, and Thought*, Cambridge, MA: Harvard University Press.

Locke, J.L., 1993, *The Child's Path to Spoken Language*, Harvard U Press, Cambridge.

Mitz, A.R., Godshalk, M., and Wise, S.P., 1991, Learning-dependent Neuronal Activity in the Premotor Cortex. Activity during the Acquisition of Conditional Motor Associations, *Journal of Neuroscience,* 11:1855-1872.

Nicod, J., 1970, *Geometry and Induction*, (Jean Bell and Michael Woods, Translators), University of California Press, Berkeley, CA.

Piattelli-Palmarini, M., 1980, *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*, Harvard University Press.

Pulvermüller, F. (2003) *The Neuroscience of Language*. Cambridge: Cambridge University Press.

Room, T.G., 1967, *A Background (Natural, Synthetic and Algebraic) to Geometry*, Cambridge: Cambridge University Press.

Ullman, M., 2004, Contributions of memory circuits to language: the declarative/procedural model, *Cognition* 92: 231-270.

Wohlschläger, Gattis, M., and Bekkering, H., 2003, Action generation and action perception in imitation: an instance of the ideomotor principle, *Phil. Trans. R. Soc. Lond.* B, 358:501–515.

Wray, A., 2000, Holistic utterances in protolanguage: the link from primates to humans. In Knight, C., Studdert-Kennedy, M. & Hurford, J. (eds.). The evolutionary emergence of language. Stanford, CA: Cambridge University Press.