Journal of Physiology - Paris 102 (2008) 4–20

Contents lists available at ScienceDirect

Journal of Physiology - Paris

journal homepage: www.elsevier.com/locate/jphysparis

From grasp to language: Embodied concepts and the challenge of abstraction

Michael A. Arbib *

Computer Science, Neuroscience and USC Brain Project, University of Southern California, Los Angeles, CA 90089-2520, USA

ARTICLE INFO

Keywords:
Grasp
Language
Embodied concepts
Abstraction
Mirror neurons
Gestural origins

ABSTRACT

The discovery of mirror neurons in the macaque monkey and the discovery of a homologous "mirror system for grasping" in Broca's area in the human brain has revived the gestural origins theory of the evolution of the human capability for language, enriching it with the suggestion that mirror neurons provide the neurological core for this evolution. However, this notion of "mirror neuron support for the transition from grasp to language" has been worked out in very different ways in the Mirror System Hypothesis model [Arbib, M.A., 2005a. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics (with commentaries and author's response). Behavioral and Brain Sciences 28, 105–167; Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. Trends in Neuroscience 21(5), 188–194] and the Embodied Concept model [Gallese, V., Lakoff, G., 2005. The brain's concepts: the role of the sensory-motor system in reason and language. Cognitive Neuropsychology 22, 455–479]. The present paper provides a critique of the latter to enrich analysis of the former, developing the role of schema theory [Arbib, M.A., 1981. Perceptual structures and distributed motor control. In: Brooks, V.B. (Ed.), Handbook of Physiology – The Nervous System II. Motor Control. American Physiological Society, pp. 1449–1480].

1. The mirror system hypothesis, briefly recalled

Any normal child reared in a human society will acquire language. Some argue that this is because Universal Grammar – the set of basic structures of the grammars of all possible human languages – is innate, so that the child need simply hear a few sentences to "set the parameter" for each key principle of the grammar of her first language (Baker, 2001; Chomsky and Lasnik, 1993). Others have argued that the modern child receives rich language stimuli within social interactions and needs no innate grammar to acquire the particular sounds (phonology) of the language, and then masters an ever increasing stock of words as well as constructions that arrange words to compound novel meanings. In either case, there is something unique about the human brain which makes it *language-ready,* in the sense that a human child can learn language while infants of other species cannot. We use a comparison of human brains with those of macaque monkeys to introduce one account of how biological evolution yielded the human language-ready brain (see also (Arbib and Bota, 2003; Deacon, 2007)).

The system of the macaque brain for visuomotor control of grasping has its premotor outpost in an area called F5 which contains a set of neurons, *mirror neurons*, such that each one is active not only when the monkey executes a specific grasp but also when the monkey observes a human or other monkey execute a more-or-less similar grasp (Rizzolatti et al., 1996). Thus macaque F5 contains a *mirror system for grasping* which employs a similar neural code for *executed* and *observed* manual actions. It is important to note that in addition, F5 contains the control of *canonical neurons* which are active for execution of grasps but not for observation of the grasps of others, and other classes of neurons as well. Canonical and mirror neurons are anatomically segregated to distinct subregions F5ab and F5c, respectively, of area F5.

The region of the human brain homologous to macaque F5 is thought to be Brodmann area 44, part of Broca's area, traditionally thought of as a speech area, but which has been shown by brain imaging studies to be active also when humans either execute or observe grasps. It is posited that the mirror system for grasping was also present in the common ancestor of humans and monkeys (perhaps 20 million years ago) and that of humans and chimpanzees (perhaps 5 million years ago). Moreover, the mirror neuron property resonates with the *parity requirement* for language – that what counts for the speaker must count approximately the same for the hearer. In addition, normal face-to-face speech involves manual and facial as well as vocal gestures, while signed languages are fully developed human languages. These findings ground "The Mirror System Hypothesis" (Arbib and Rizzolatti, 1997; Rizzolatti and Arbib, 1998): *The* parity requirement for language in humans is met because Broca's area evolved atop the mirror system for grasping which provides the capacity to generate and recognize a set of actions.

* Tel.: +1 213 740 9220.
E-mail address: arbib@pollux.usc.edu

0928-4257/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jphysparis.2008.03.001

In putting parity at stage center in this account, we adhere to the view that the primary function of language is communication. Others have espoused the alternative view that language evolution could have obeyed an adaptive pressure for developing higher cognitive abilities and that verbal communication would be a secondary benefit. I have two comments. (i) Language is a shared medium, and thus parity is essential to it. No matter how useful a word may be as a tool for cognition, we must learn the word in the first place; and we must then engage in numerous conversations if, in concert with our own thoughts, we are to enrich our understanding of any associated concept and our ability to make fruitful use of it. (ii) Having said this, I readily admit, as is clear from the preceding, that language is a powerful tool for thought (though much thought is non-verbal). Thus, while I believe that parity was the key to getting language (or, more strictly, protolanguage – see below) "off the ground", both the external social uses of language and the internal cognitive uses of language could have provided powerful and varied adaptive pressures for further evolution of such capacities as anticipation, working memory, and autobiographic memory as language enriched both our ability to plan ahead, explicitly considering counter-factual possibilities, and mulling over past experience to extract general lessons. Indeed, where we lay stress on parity in the evolution of the language-ready brain, Aboitiz et al. (Aboitiz, 1995; Aboitiz et al., 2006; Aboitiz and Garcia, 1997) lay primary stress on the evolution of working memory systems. I see such alternatives as complementary, rather than either excluding the other.

With this, let me turn to a fuller exposition of the "Mirror System Hypothesis". I start with a few comparative comments concerning imitation to introduce key differences between monkey, ape and human that are relevant to understanding what such evolution may have involved. Monkeys have, at best, a very limited capacity for imitation (Visalberghi and Fragaszy, 1990; Voelkl and Huber, 2007), far overshadowed by what I call *simple imitation* as exhibited by apes. Myowa-Yamakoshi and Matsuzawa (1999) observed that chimpanzees took 12 or so trials to learn to "imitate" a behavior in a laboratory setting, focusing on bringing an object into relationship with another object or the body, rather than the actual movements involved. Byrne and Byrne (1993) found that gorillas learn complex feeding strategies but may take months to do so. Consider eating nettle leaves. Skilled gorillas grasp the stem firmly, strip off leaves, remove petioles bimanually, fold leaves over the thumb, pop the bundle into the mouth, and eat. The challenge for acquiring such skills is compounded because ape mothers seldom if ever correct and instruct their young (Tomasello, 1999) and because the sequence of "atomic actions" varies greatly from trial to trial. Byrne (2003) implicates *imitation by behavior parsing*, a protracted form of statistical learning whereby certain *subgoals* (e.g., nettles folded over the thumb) become evident from repeated observation as being common to most performances. In his account, the young ape may acquire the skill over many months by coming to recognize the relevant subgoals and derive action strategies for achieving them by trial-and-error.

This ability to learn the overall structure of a specific feeding behavior over many, many observations is very different from the human ability to understand any sentence of an open-ended set as it is heard, and generate another novel sentence as an appropriate reply. In many cases of praxis (i.e., skilled interaction with objects), humans need just a few trials to make sense of a relatively complex behavior if the constituent actions are familiar and the subgoals these actions must achieve are readily discernible, and they can use this perception to repeat the behavior under changing circumstances. We call this ability *complex imitation* (extending the definition of (Arbib, 2002) to incorporate the goal-directed imitation of Wohlschläger et al. (2003)). With such considerations in mind, I have elaborated the "Mirror System Hypothesis" (see (Arbib, 2005a) for a review, and commentaries on current controversies), defining an evolutionary progression of seven stages, S1 through S7:

- **S1:** Cortical control of hand movements.
- **S2:** A mirror system for grasping, shared with the common ancestor of human and monkey.

I stress that a mirror system does not provide imitation in itself. A monkey with an action in its repertoire may have mirror neurons active both when executing and observing that action yet does not repeat the observed action. Nor, crucially, does it use observation of a novel action to add that action to its repertoire. Thus, we hypothesize that evolution embeds a monkey-like mirror system in more powerful systems in the next two stages.

- **S3:** A simple imitation system for grasping, shared with the common ancestor of human and apes.
- **S4:** A complex imitation system for grasping which developed in the hominim line since that ancestor.

Each of these changes can be of evolutionary advantage in supporting the transfer of novel skills between the members of a community, involving praxis rather than explicit communication. We now explore the stages whereby our distant ancestors made the transition to *protolanguage*, in the sense of a communication system that supports the ready addition of new utterances by a group through some combination of innovation and social learning – it is open to the addition of new "protowords", in contrast to the closed set of calls of a group of nonhuman primates – yet lacks any tools, beyond mere juxtaposition of two or three protowords, to put protowords together to continually create novel utterances from occasion to occasion. Arbib et al. (submitted for publication), summarizing data on primate communication, note that monkey vocalizations are innately specified (though occasions for using a call may change with experience), whereas a group of apes may communicate with novel gestures, perhaps acquired by *ontogenetic ritualization* (Tomasello et al., 1997) whereby increasingly abbreviated and conventionalized form of an action may come to stand in for that action, an example being a beckoning gesture recognized by the child as standing for the parent's action of reaching out to grasp the child and pull it closer. This supports the hypothesis that it was gesture rather than vocalization (Seyfarth et al., 2005) that created the opening for greatly expanded gestural communication once complex imitation had evolved for practical manual skills. The expanded version of the "Mirror System Hypothesis" addresses this by positing the next two stages to be:

- **S5:** *Protosign*, a manual-based communication system breaking through the fixed repertoire of primate vocalizations to yield an open repertoire.
- **S6:** *Protolanguage as Protosign and Protospeech*: an expanding spiral of conventionalized manual, facial and vocal communicative gestures.

The transition from complex imitation and the small repertoires of ape gestures (perhaps 10 or so novel gestures shared by a group) to protosign involves, in more detail, first pantomime of grasping and manual praxic actions then of non-manual actions (e.g., flapping the arms to mime the wings of a flying bird), complemented by conventional gestures that simplify, disambiguate (e.g., to distinguish "bird" from "flying") or extend pantomime.

Pantomime transcends the slow accretion of manual gestures by ontogenetic ritualization, providing an "open semantics" for a large set of novel meanings (Stokoe, 2001). However, such pantomime is inefficient – both in the time taken to produce it, and in

the likelihood of misunderstanding. Conventionalized signs extend and exploit more efficiently the semantic richness opened up by pantomime. Processes like ontogenetic ritualization can convert elaborate pantomimes into a conventionalized "shorthand", just as they do for praxic actions. This capability for protosign – rather than elaborations intrinsic to the core vocalization systems – may then have provided the essential scaffolding for protospeech and evolution of the human language-ready brain. Arbib (2005b) suggest how this might have come about, while MacNeilage and Davis (2005) offer a counter-argument. Jürgens (1979, 2002) provides the relevant neurobiological data, though working primarily with squirrel monkey rather than macaques. He found that voluntary control over the initiation and suppression of monkey vocalizations relies on the mediofrontal cortex including anterior cingulate gyrus – but note that this is initiation and suppression of calls from a small repertoire, not the dynamic assemblage and co-articulation of articulatory gestures that constitutes speech. Such findings suggest that the anterior cingulate cortex is involved in the volitional initiation of monkey vocalization. Thus a major achievement of the "Mirror System Hypothesis" is to develop a plausible explanation as to why Broca's area corresponds to F5 rather than the vocalization area of cingulate cortex by showing how manual gesture and pantomime could ground protospeech via protosign. Ferrari et al. (2003, 2005) found that F5 mirror neurons include some for oro-facial gestures involved in feeding. Moreover, some of these gestures (such as lip-smack and teeth chatter) do have auditory side-effects which can be exploited for communication. This system has interesting implications for language evolution (Fogassi and Ferrari, 2004), but is a long way from mirror neurons for speech. Intriguingly, squirrel monkey F5 does have connections to the vocal folds (Jürgens, personal communication, 2006), but these are solely for closing them and are not involved in vocalization (but see Coudé et al., 2007). We thus extend the argument in Arbib (2005b) by hypothesizing that the emergence of protospeech on the scaffolding of protosign involved expansion of the F5 projection to the vocal folds to allow for vocalization to be controlled in coordination with the control of the use of tongue and lips as part of the ingestive system.

We now come to the final stage, the transition from protolanguage to language:

- **S7:** *Language*: the development of syntax and compositional semantics.

This may have involved grammatically specific biological evolution. Pinker and Bloom (1990) argue that Universal Grammar is innate, evolving through multiple stages, but their definition of universal grammar is incomplete, and some of their stages seem as amenable to cultural as to biological evolution. However, I am among those who argue that the diversity of grammar is to be captured in the history of different societies rather than in the diversity of the genes. The nature of the transition to language remains hotly debated.

Although EMSH, the particular Extension of the "Mirror System Hypothesis" presented here, posited that complex imitation evolved first to support the transfer of praxic skills and then came to support protolanguage, it is important to note its crucial relevance to modern-day language acquisition and adult language use. Complex imitation has two parts: (i) the ability to perceive that a novel action may be approximated by a composite of known actions associated with appropriate subgoals; and (ii) the ability to employ this perception to perform an approximation to the observed action, which may then be refined through practice. Both parts come into play when the child is learning a language whereas the former predominates in adult use of language as the emphasis shifts from mastering novel words and constructions to finding the appropriate way to continue a dialogue.

Another influential account has been given by Deacon (1997) who, as we do, sees language function as supported by many evolutionary modifications of the brain rather than one "big bang mutation", though he gives primacy to symbolic reference where we emphasize the way in which language may have built on new, communicative uses for brain mechanisms evolved for *praxis*, practical interactions with objects. He has invoked the Baldwin Effect (Baldwin, 1896) (in which behavioral plasticity enabling the production of acquired adaptations serves as an evolutionary precursor to a more innately grounded analogue of this adaptation) to support the evolution of biases and aids to learning language, without requiring the replacement of learned with innate knowledge of syntax such as postulated for Universal Grammar. More recently, he has revisited this argument in terms of niche construction (Deacon, 2003), suggesting how persistent socially maintained language use might be understood as a human-constructed niche that exerts significant selection pressures on the organism to adapt to its functional requirements. In other words, languages evolved to be learnable by humans at the same time as human brains evolved to be better suited to learn language. This approach assumes that language-like communication was present in some form for an extensive period of human prehistory. EMSH emphasizes the specific roles of protosign and protospeech, rather than language per se.

Sternberg and Christiansen (2006) argue, as does Deacon, that languages have evolved to fit preexisting learning mechanisms, noting that sequential learning is one possible contender since sequential learning and language both involve the extraction and further processing of elements occurring in temporal sequences (see, e.g., (Dominey and Hoen, 2006) for an argument based on computational modeling as well as neuroimaging). They note that human sequential learning appears to be more complex (e.g., involving hierarchical learning) than that which has been observed in non-human primates – this would accord with our emphasis on complex imitation.

Work on the "Mirror System Hypothesis" to date falls short of explaining the transition from protolanguage to language per se. As noted earlier, I reject the view that Universal Grammar establishes within the infant brain a range of parameters such that the child acquires the syntax of its native language by setting each parameter simply by hearing a few sentences to determine which value of the parameter is consistent with them (Chomsky and Lasnik, 1993; Lightfoot, 2006; Arbib, 2007, gives a critique). Rather, I agree with the construction grammarians (Goldberg, 2003) who see each language as defined by an idiosyncratic set of constructions which combine *form* (how to aggregate words) with *meaning* (how the meaning of the words constrains the meaning of the whole). Various authors (Arbib and Hill, 1988; Tomasello, 2003) have explained how modern children acquire words and constructions without invoking Universal Grammar. Hill (1983) showed that the child may first acquire what the adult perceives as two-word utterances as holophrases (e.g., "want-milk") prior to developing a more general construction (e.g., want x") in which "x" can be replaced by the name of any "wantable thing". Further experience will yield more subtle constructions and the development of word classes like "noun" defined by their syntactic roles in a range of constructions rather than their meaning.

Ontogeny does not in this case recapitulate ontogeny. Adult hunters and gatherers had to communicate about situations outside the range of a modern 2-year-old, and protohumans were not communicating with adults who already used a large lexicon and set of constructions to generate complex sentences. Nonetheless, I argue that protolanguage and language emerged through the invention of an increasingly subtle interweaving of (proto)words and (proto)constructions. We should not put our faith in Universal Grammar but rather seek to identify the hitherto Unidentified

Gadgets (a coinage of Jean-Roger Vergnaud) that make human use of language possible. I suggest that the same basic mechanisms may have served both protohumans inventing language and modern children acquiring the existing language of their community (Arbib, 2008):

1. The ability to create a novel gesture or vocalization and associate it with a communicative goal.
2. The ability both to perform and perceive such a gesture or vocalization would improve with experience as its use spread within the community, as would sharpening of the perception of occasions of use by members of the community.
3. Commonalities between two structures could yield to the isolation of that commonality as a gesture or vocalization betokening some shared aspect of the event, object or action denoted by each of the two structures (see (Wray, 2000) for how this might have operated in protohumans; and (Kirby, 2000) for a related computer model). This could in time lead to the emergence of a construction for "putting the pieces back together", with the original pieces becoming instances of an ever wider class of slot fillers. It is the ability for complex imitation that makes these processes possible. In the case of protohumans, this could lead to the invention of new (proto)words and constructions. In the case of the modern child, it provides the basis for understanding that strings of sounds can be dissected into strings of words, that these words can be grouped by constructions. The constructions become of greater or more focused applicability both on a historical time-scale as new words and constructions are invented over the course of many generations; and on a developmental time-scale as the child has more experience of using fragments of the ambient language to understand and be understood.

2. Embodied concepts

One can agree with the general claim that the mirror system for grasping grounded the brain's capability for language without accepting the particular evolutionary stages outlined in the previous section. Analysis of other attempts to work out the general insight may thus yield a deeper analysis of the pros and cons of EMSH. To this end, the rest of this paper develops a constructive critique of the paper of Gallese and Lakoff (2005), henceforth referred to as G&L, which combines the insights of Vittorio Gallese, one of the neurophysiologists on the team which discovered mirror neurons for grasping in area F5 of the macaque (di Pellegrino et al., 1992), and the linguist George Lakoff who is particularly well known for his insights into how much metaphor contributes to the creation of meaning in language (Lakoff and Johnson, 1980). The key notion for Gallese and Lakoff (2005), who reject the arguments of "early cognitivism, [namely that] concepts are symbolic representations by nature, and as thinking, they can be reduced to symbolic (not neural) computation" is that "conceptual knowledge is *embodied*, that is, it is mapped within our sensory-motor system [my italics]." They advance their argument and link it to language through three claims:

(a) Imagining and doing use a shared neural substrate.[1] Here, imagining is taken to be a *mental simulation* of action or perception, using many of the same neurons as in actual acting or perceiving (Gallese, 2003) (see (Goldman, 2006) for more on this theme; and (Jacob and Jeannerod, 2005) for a critique of this rather broad use of the term *simulation*.)

(b) Understanding *is* imagination so that what you understand of a sentence in a context is the meaning of that sentence in that context. If you cannot imagine picking up a glass or seeing someone picking up a glass, then you cannot understand the sentence "Harry picked up the glass."

(c) Imagination, like perceiving and doing, is *embodied*, that is, structured by our constant encounter and interaction with the world via our bodies and brains.

My goal in this paper is to examine these claims and offer what I consider to be a firmer foundation for future research on the extension of sensorimotor processes to support the functions of language. Here are initial responses that will be developed in what follows:

For (a): The idea within the modern psychological literature of cognition as in some sense involving simulation goes back at least to Craik (1943) and was vigorously developed by Gregory (1969), MacKay (1966) and Minsky (1965) in the 1960s. However, G&L (see also Gallese and Goldman, 1998) seem to deny a possible distinction between, for example, the actual circuitry used to execute an action and the (neurally realized) model used to plan that action.

Among the Oxford English Dictionary's definitions of simulation, three seem pertinent here:

(1b) Tendency to assume a form resembling that of something else; unconscious imitation.

(2) A false assumption or display, a surface resemblance or imitation, *of* something.

(3) The technique of imitating the behavior of some situation or process (whether economic, military, mechanical, etc.) by means of a suitably analogous situation or apparatus, etc.

It seems that G&L assume an unusual variant of (1b) and (2), in which *surface resemblance* is replaced by *neural activity*, arguing that one person simulates the behavior of another if their neural activity resembles that which they would generate were they to behave in the same way. By contrast, (3) (the version which is natural for me in my role as a computational neuroscientist, using computers to simulate brains) does not require a congruence of bodily and neural structure but rather requires that data on measurements of the original system be matched with some degree of precision by the results generated by running the model with appropriate input data. Thus, while simulation may, as envisioned by G&L, in some cases involve activation of the relevant motor circuitry (but with motor output inhibited) to estimate, e.g., the effort or outcome of the given action, simulation at this level would not be appropriate when planning, for example, a three week vacation.

G&L to some extent anticipate this concern when they distinguish three categories of the *canonical* neurons of area F5ab:

- Firing of general-purpose neurons indicates the general goal of the action (e.g., grasp, hold, tear an object), not how the action is carried out.
- Firing of *manner* neurons correlates with the various ways in which a particular action can be executed (e.g., grasping an object with the index finger and the thumb, but not with the whole hand), while
- *Phase* neurons deal with the temporal phases into which purposeful actions are segmented (e.g., hand/mouth opening phase, or hand/mouth closure phase).

G&L comment that the general-purpose neurons can never function alone in action, since all actions are carried out in some manner and are in one phase or another at some time. G&L see it

[1] "Imagining and doing use a shared neural substrate" is trivially true if we take that neural substrate to be the whole human brain. But the more focused claim "The minimal neural system in the human brain that supports 'doing' overlaps the minimal neural system in the human brain that supports 'acting'" is indeed true, and so we will interpret (a) in the latter sense.

as strong prima facie evidence for simulation that when a monkey attends to a graspable object, only the neurons with the right manner of grasping for that object fire (see Gallese, 2003). They then speculate that the general-purpose neurons might fire without a manner subcluster firing, *in simulation*. That is, one should be able to simulate in imagination carrying out a *general* action without specifying manner. However, we shall see below (in discussing the FARS model) that planning an action may involve far more than observing an object, so that the neural activity required for imagination may place F5 activity within a much larger complex. Below, I return to the fact that G&L here implicate canonical neurons in their account of simulation. They also implicate mirror neurons, as we shall also see below.

For (b): It seems to me a dangerous move to equate "what you understand of a sentence in a context" with "the meaning of that sentence in that context", in that it makes meaning not only subjective but ever changeable. Consider *A* saying to *B*, "I don't understand what you mean", when *B* has just uttered a sentence *X*, and then saying "Oh, I get it" after some ensuing dialogue. The equation of meaning with understanding would imply that *X* was initially meaningless. Perhaps the matter is resolved if we adopt a notation like M(*X*,*A*,*t*) to denote the meaning of sentence *X* to person *A* at time t. Perhaps "the" meaning M(*X*) of *X* would be the value of M(*X*,*A*,t) common to most speakers *A* of a community within some suitable range of times *t*, if such a common value exists. *A* question of major concern to us, then, is to determine what it is about 2 people that ensures that for a wide range of utterances, the utterance of a sentence *X* will be generated by a speaker with the intention to convey M(*X*) and interpreted by the hearer to convey that same meaning M(*X*). Perhaps an even more appropriate notion would be M(*X*,*C*), the meaning of *X* in context *C*, which only equals M(*X*,A,t) if *A* recognizes at time *t* that *C* is the context relevant to understanding *X*.

Consider the claim "If you cannot imagine picking up a glass or seeing someone picking up a glass, then you cannot understand that sentence 'Harry picked up the glass.'" This claim seems eminently plausible for a sentence describing a concrete interaction of a person with an object. But what about a truly abstract sentence? Should we assert "If you cannot imagine resolving the matter, and you cannot imagine adopting a notation, and you cannot imagine denoting the meaning of a sentence *X*, and you cannot imagine being person *A*, and you cannot imagine being at time *t*, then you cannot understand that sentence 'Perhaps the matter is resolved if we adopt a notation like M(*X*,*A*,*t*) to denote the meaning of sentence *X* to person *A* at time *t*'"? G&L offer a step in the right direction when they show how metaphor can extend our understanding of events in one domain to events that might otherwise be hard to understand in another domain. However, this can only be part of our ability to deal with situations of ever increasing abstraction. I cannot pretend to offer a comprehensive account of how this occurred, but will offer some preliminary comments in the discussion of infidelity in the section "Towards Abstraction".

For (c): We can imagine more than we can do with the bodies we have. For example, we can imagine flying through the winds of Jupiter or finding a proof of Fermat's last theorem but (with very, very few exceptions in the latter case) we can do neither. The former is interesting in that it does rest on embodiment, but an embodiment that is not ours – the amazing power of language is that it makes the counterfactual understandable. The latter is interesting in that – as in the above paragraph – it makes clear that our development as language users allows us to understand sentences for which embodiment is not relevant. All that is required is the ability to form strings of symbols and recognize the relations between them.

I share with G&L the view that "sensory-motor brain mechanisms [were adapted] to serve new roles in reason and language, while retaining their original functions as well." But what of their claim that

- Language exploits the pre-existing sensorimotor-modal character of the sensory-motor system.
- there is no single "module" for language [cf. Arbib, 1987],
- human language makes use of mechanisms also present in non-human primates?

I believe that what Gallese & Lakoff mean by this is that the sensory motor-system of the human brain is "pre-existing" in the sense that it adheres to a generic structure shared with other primates and their common ancestor, and that this in itself supports language. This then begs the question of why humans can acquire language and other primates cannot. Would they claim it holds just because, e.g., this generic system is bigger in humans? By contrast, I would disambiguate the claim by the following rephrasing: "The brain mechanisms which make the human use of language possible *include* structures which are highly similar to the sensory-motor system of other primates, and there is no separate 'module' for language which is strictly separable from these mechanisms." EMSH is a working out of this latter meaning on an evolutionary time scale, in which circuitry in ancestral primates has been duplicated then specialized in various ways so that human brains are not just bigger macaque or chimpanzee brains but possess unique properties that make the use of language possible.

In an earlier critique, Mahon and Caramazza (2005) focus on the fact that G&L seem to assume that motor production circuitry is *necessarily* involved in the recognition of visually presented actions. They call this proposal the *Motor Theory of Action Recognition*, echoing the *Motor Theory of Speech Perception* (Liberman et al., 1967) which holds that the listener recognizes speech by activating the motor programs that would produce sounds like those that are being heard. Mahon and Caramazza (2005) take a rather different tack from that developed in the present article. They review the performance of patients with apraxia, an impairment in using objects that cannot be attributed to aphasia, sensory impairment, or an impairment to basic motor responses. The motor theory of action recognition predicts that apraxia will necessarily be associated with an inability to correctly recognize visually presented actions. Mahon and Caramazza (2005) review contrary data on patients who are impaired in using objects but not impaired at distinguishing correct from incorrect object-associated actions. In fact, just such dissociations were advanced by Barrett et al. (2005) as an argument against the "Mirror System Hypothesis", and I was at pains in (Arbib, 2006) to show that such dissociations do not militate against the "Mirror System Hypothesis". Rather, the "Mirror System Hypothesis" posits that mechanisms that support language in the human brain evolved atop the mirror system for grasping but – in distinction from G&L, we may now note – is not restricted

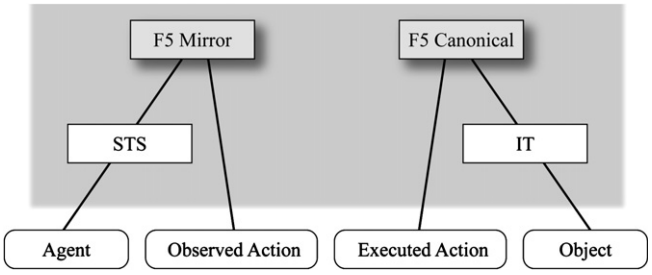

Fig. 1. A schematic of the canonical and mirror neurons stressing that executing an action requires linking that action to the goal object (as recognized, e.g., by IT), while recognizing an action executed by another requires the further linkage of the action to the agent (who might be recognized by STS) (Arbib and Mundhenk, 2005).

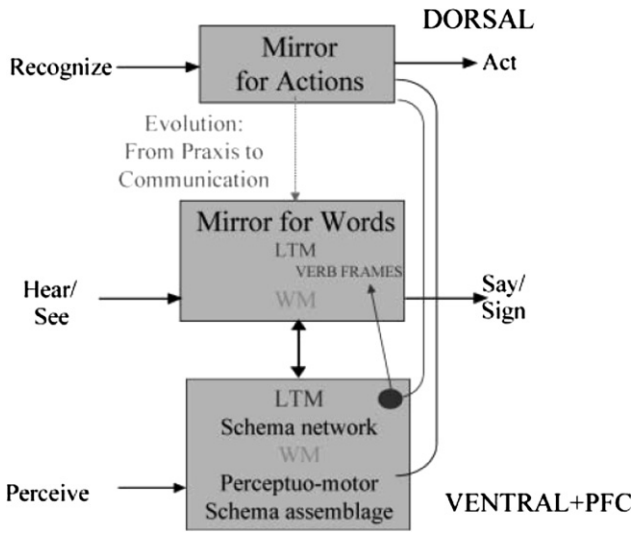

Fig. 2. Words link to schemas, not directly to the dorsal path for actions (Arbib, 2006).

to the circuits of the ancestral sensorimotor system. (Fig. 2 summarizes the key structures envisaged in Arbib (2006).) However, when Mahon and Caramazza (2005) say "To the degree that one is not compelled to infer the necessity of muscle activation for conceptual understanding, why then make the inference that the activation of motor cortices is necessary for understanding?" they may be failing to understand the importance of embodiment in grounding much of our conceptual repertoire. There is a large spectrum of circumstances – including sensorimotor expectations – associated with an action, and these enrich a concept far beyond dictionary definitions.

3. Schema theory, briefly recalled

To put G&L's claims in perspective, I need to recall my version of schema theory (Arbib, 1981), henceforth denoted ST. This represents knowledge actively, with schemas (like neural networks) serving as both repositories and processors of information. Indeed, information in the brain/schema network does not exist apart from the processes for using it. ST offers an *action-oriented view* which emphasizes that perceptual acts take place against a background; that we cannot recognize details in a void; and that present schemas strongly color the choice of each addition. We always act within some context. *Perceptual schemas* serve perceptual analysis, while *motor schemas* provide control systems for movement. ST stresses that the organism perceives and acts within an *action-perception cycle* in which actions guide perception as much as perception guides action.

Why not combine perceptual and motor schemas into a single notion of schema that integrates sensory analysis with motor control? Certainly, there are cases where such a combination makes sense. However, recognizing an object may be linked to many different courses of action. Recognizing an apple, for example, one might place it in one's shopping basket; place it in a bowl; pick it up; peel it; cook with it; eat it; discard it if it is rotten, etc. Of course, once one has decided on a particular course of action then specific perceptual and motor subschemas must be invoked. But note that, in the list just given, some items are apple-specific whereas others invoke generic schemas for reaching and grasping. It was considerations like this that underwrote the separation of perceptual and motor schemas (Arbib, 1981) – a given action may be invoked in a wide variety of circumstances; a given perception may precede many courses of action. There is no one grand "apple schema" which links all "apple perception strategies" to "every action that involves an apple". Moreover, in the schema-theoretic approach, "apple perception" is not mere categorization "this is an apple" but may provide access to a range of parameters relevant to interaction with the apple at hand. Again, peeling a banana is a distinct skill from the one skill of peeling an apple or pear. But we can recognize a banana without mentally peeling it – we might, e.g., go directly from visual appearance to the anticipation of taste on eating it.

The overall schema network develops as new schemas are formed both de novo by trial-and-error and from old schemas whether by combining schemas in *coordinated control programs* and *schema assemblages,* or by *abstracting* from sets of extant schemas. In other words, knowledge is grounded in the embodied organism's interaction with its world, but *not limited to it.* Human cognition is grounded in sensorimotor schemas which mediate our embodied interactions with the world, but these make it possible to acquire abstract schemas whose processing may link to embodiment only via its development within the schema network (a view which, apart from the neural trappings, is most associated with the work of the Swiss genetic epistemologist, Jean Piaget – see, e.g., Piaget, 1954). However, once it has developed, an abstract schema may also be able to conduct its processing in relative isolation from this background. Thus, when a mathematician proves a theorem about an abstract space, he may exploit his intuitions of the everyday space of embodied interaction, or may follow paths of abstract reasoning powerful enough to establish general results that contradict those intuitions of everyday experience.

But how do schemas relate to neural circuitry? Although analysis of some overall function may require no hypotheses on the localization of constituent schemas, the schemas can be linked to a structural analysis as and when this becomes appropriate. In general a given schema ("functional module") may be subserved by the interaction of several "(structural) brain modules", and a given "brain module" may be involved in subserving a number of different functions (Mountcastle, 1978). Given hypotheses about the neural localization of schemas, we may then model a brain region to see if its known neural circuitry can be shown to implement the posited schema. When the model involves properties of the circuitry that have not yet been tested, it lays the ground for new experiments.

A schema may be instantiated to form multiple *schema instances.* e.g., given a schema that represents generic knowledge about some object, several instances of the schema may be activated to subserve our perception of a scene containing several such objects. Schema instances can become *activated* in response to certain patterns of input from sensory stimuli (data driven) or other schema instances that are already active (hypothesis driven). The *activity level* of an instance of a perceptual schema represents a "confidence level" that the object represented by the schema is indeed present; while that of a motor schema may signal its "degree of readiness" to control some course of action. Schemas are the "programs" for *cooperative computation*, based on the competition and cooperation of concurrently active schema instances. *Cooperation* yields a pattern of "strengthened alliances" between mutually consistent schema instances that allows them to achieve high activity levels to constitute the overall solution of a problem (as perceptual schemas become part of the current short-term model of the environment, or motor schemas contribute to the current course of action). It is as a result of *competition* that instances which do not meet the evolving consensus lose activity, and thus are not part of this solution (though their continuing subthreshold activity may well affect later behavior). A corollary to this view is that knowledge receives a distributed representation in the brain. A multiplicity of different representations must be linked into an

integrated whole, but such linkage may be mediated by distributed processes of competition and cooperation. There is no one place in the brain where an integrated representation of space, for example, plays the sole executive role in linking perception of the current environment to action.

4. Embodied concepts and parameters

A cat has three gaits – strutting, trotting, and galloping – which correspond to low, intermediate and high speed locomotion and the neural firing that encodes it (Grillner and Wallen, 2002; Tomonaga et al., 2004). G&L call this firing a *neural parameter,* with low, medium, and high firing frequencies as the *values* of that parameter. However, they fail to note that the *quantitative* variation of a neural parameter in many cases fails to yield a discrete set of *qualitative* categories – just consider the neural firing that controls eye movement or finger aperture. For another example, there is a large range of "peripersonal space on my left" that I can reach with my right hand as well as my left – there is no motoric division between right and left. Thus, some conceptual distinctions arise from motor discontinuities, but others are imposed top-down by words serving to anchor conventionalized distinctions. Again, G&L note that if the force required to push an object is very high, what is required is *shoving* rather than mere *pushing*. Shoving requires a different motor program: setting the weight on the back foot, and so on. They then assert that "the choice of parameter values also determines motor programs for humans as well as for cats. Moreover, parameter values govern simulations as well. Imagining pushing is different from imagining shoving." G&L make the claim

> Parameters can be seen as "higher-level" features of neural organization, while the neural firings in particular motor circuits for various gaits can be seen as being at a "lower level" of organization. To the higher-level parameters, the lower-level structure is "invisible." Parameterization thus imposes a hierarchical structure on the neural system.

However, G&L do note a number of parameters of action – such as applied force, direction of motion – for which there are neural correlates ("neural parameters") but for which there is no separation of a continuous range into subranges with qualitatively distinct correlates. Yet we may still invent words to arbitrarily suggest ranges of parameter values. A light force might be described as "a feather touch". I would see this as an argument against *mere* sensorimotor grounding – these categories are verbal overlays, rather than inherent in the action itself. My point is not to deny that many concepts have an important sensorimotor grounding. It is, rather, to suggest that G&L over-emphasize it to the point of excluding what is distinctive about the conceptual apparatus of the human brain. The skill of a sportscaster describing what he sees is different from the skill of the athlete he observes. Thus the parameters involved in describing a skill are not the automatic readout of the neural code for that skill. Moreover, we can imagine ourselves flying but have no neural parameters for wing control and our language contains verbs for far more than overt actions. Conversely, animals (like the locomoting cat) have parameters for action but no way to communicate about them other than as a side-effect of the action itself.

G&L assert that "Parameters and their values are accessible to consciousness, while anything below the parameter value level is inaccessible." Here, it seems that G&L and ST use the term "parameter" in different ways. For ST, the parameters which control a hand action may include the size, shape and location of the object to be grasped, rather than mere qualitative descriptors. The encoding of such parameters via the dorsal pathway (the "how" pathway) is inaccessible to consciousness. Rather, it seems that it is the ventral pathway (the "what" pathway) that provides a separate encoding related to conscious decision-making or description (Bridgeman, 2005; Goodale and Milner, 1992; Jeannerod et al., 1994). There is nothing that enforces an isomorphism, neural or otherwise, between concepts and verbal expression. Of course, a similar stricture applies to concepts which do not involve actions, like "dog", "blue", "neural" and "improbably".

G&L use the concept *grasp* to illustrate their ideas on embodied concepts:

1. *Information structure*: they note that the information structure needed to characterize the conceptual structure of *grasp* is available at the neural level in the sensory-motor system.
2. *Multimodality:* G&L see mirror neurons and other classes of premotor and parietal neurons as inherently "multimodal" because the firing of a single neuron may correlate with both *seeing* and *performing* grasping. This is true, but the term "multimodality" is used more often to refer to multiple sensory modes. However, some mirror neurons are multimodal in the sensory sense. Kohler et al. (2002) found mirror neurons that respond to the sound as well as the sight of specific types of action that, like breaking a peanut, have their own distinctive sounds.
3. *Functional clusters:* G&L assert that Sensory-Motor Parity is realized in the brain through *functional clusters*, including parallel parietal-premotor networks, which form *high-level units* – characterizing the discreteness, high-level structure, and internal relational structure required by concepts.
4. *Simulation:* for G&L, any conceptualization of grasping involves simulation using the same functional clusters as used in the action and perception of grasping.
5. *Parameters:* all actions, perceptions, and simulations make use of neural parameters and their values. For example, the action of *reaching* for an object makes use of the neural parameter of direction; the action of *grasping* an object makes use of the neural parameter of force. G&L claim that such neural parameterization is pervasive and imposes a *hierarchical structure* on the brain, but ST reminds us that continuous variation in such parameters may be the key to sensorimotor coordination (with parameters "passed" from perceptual to motor schemas), with only limited cases of hierarchical structuring being related to qualitative differences in motor execution. A further point is that our imagination need not specify parameters at all, or at most need provide some coarse estimate of the parameter rather than the value we would need for successful action. I believe I understand the sentence "William Tell shot the apple on his son's head," remembering that the shooting was with bow and arrow. However, in imagining the scene, I may conjure up a mental image of the arrow quivering in the apple above the boy's head rather than simulating the action that got it there. Indeed, unless pushed to do so by a questioner, I might not have imagined whether the bow was a crossbow or a longbow – and in either case, I do not have the motor schemas for using a bow with even a modicum of skill. The "what sort of bow" critique might be addressed by G&L by invoking the firing of *general-purpose* neurons (as distinct from *manner* neurons *or phase* neurons) but none of the neurons related to shooting the arrow would help us understand the true meaning of the sentence – namely, that what was important about the action was the risk to which William Tell was forced to expose his son.
6. *Structured neural computation:* G&L state that the *neural theory of language* (Feldman and Narayanan, 2004; Lakoff and Johnson, 1999) provides a theory of neural computation in which the same neural structures that allow for movement and perception in real time and in real sensory-motor contexts also permit real-time context-based inferences in reasoning. G&L argue that "from the structured connectionism perspective, the inferential

structure of concepts is a consequence of the network structure of the brain and its organization in terms of functional clusters. This brain organization is, in turn, a consequence of ... the way in which our brains, and the brains of our evolutionary ancestors, have been shaped by bodily interactions in the world." However, a word of warning is in order. The term *neural theory of language* here refers to a particular computational model developed by Feldman & Narayan and their colleagues, and makes essentially no contact with neuroscientific data. It employs "structured neural computation" in a technical sense from the study of *artificial* neural networks This uses a structured representation of propositions in which the separate items of a logical form are mapped to separate groups of neurons in the artificial neural network (Shastri and Ajjanagadde, 1993). This provides an intriguing approach to how reasoning might be implemented on an artificial neural network but has no proven relevance to analyzing the biological networks of the human brain. Linking back to our discussion of parameters, note that, since there is much overhead in working out the details of motion, a reduced representation which is more symbol-like than simulation-like could in many cases provide a better medium for reasoning than would the use of the neural structures that "allow for movement and perception in real time and in real sensory-motor contexts." Indeed, the ability to transfer between such different representations may be the secret of human intelligence – with "motor simulation" just one among several strategies.

5. Modeling the control and recognition of grasp

Fagg and Arbib (1998) developed the FARS (Fagg–Arbib–Rizzolatti–Sakata) model for the control of *canonical neurons*, the neurons in F5 active for execution of grasps but not for observation of the grasps of others. Parietal area cIPS transforms visual input to extract the position and orientation of an object's surfaces. This information is transmitted to area AIP which extracts the *affordances* the object offers for grasping (i.e., the visually grounded encoding of "motor opportunities" for grasping the object, rather than its classification). The basic pathway AIP→F5canonical→F1 (primary motor cortex, also known as M1) of the FARS model then transforms the (neural code for) the affordance to the appropriate motor schema (F5) and thence to the appropriate detailed descending motor control signals (F1). Going beyond the empirical data then available, Fagg and Arbib (1998) stressed that in general, even when attention is focused on a single object, there may be several ways to grasp that object. The original FARS model thus hypothesized that

(a) object recognition (mediated by inferotemporal cortex IT) can influence the computation of working memory and task constraints and the effect of instruction stimuli in various areas of prefrontal cortex (PFC), and
(b) strong connections between PFC and F5 provide the data for F5 to choose one affordance from the possibilities offered by AIP.

However, contra (b), anatomical evidence (reviewed by Rizzolatti and Luppino, 2001) was later found that demonstrated that connections from PFC to F5 in macaque are very limited whereas rich connections exist between prefrontal cortex and AIP. Furthermore AIP, unlike F5, receives direct input from IT (Webster et al., 1994). Rizzolatti and Luppino (2003) then suggested that FARS be modified so that information on object semantics and the goals of the individual directly influence AIP rather than F5. Thus, selection of an appropriate grip would occur in AIP by biasing those affordances that would lead to the grip appropriate to the individual's current intentions. The affordance selected in AIP then establishes in the F5 neurons a command which reaches threshold for the appropriate grip once it receive a "go signal" from F6 (pre-SMA) which (in concert with the basal ganglia) will determine whether external and/or internal contingencies allow the action execution.

Just as we have embedded the F5 canonical neurons in a larger system involving both the parietal area AIP and the inferotemporal area IT, so do we now stress that the F5 mirror neurons are part of a larger mirror system that includes (at least) parts of the superior temporal gyrus (STS) and area PF of the parietal lobe. We now discuss a model of this larger system, the MNS model (Oztop and Arbib, 2002). One path in this model corresponds to the basic pathway AIP→F5canonical→M1 of the FARS model (but MNS does not include the prefrontal influences). Another pathway (MIP/LIP/VIP→F4) completes the "canonical" portion of the MNS model, with intraparietal areas MIP/LIP/VIP providing object location information which enables F4 to instruct F1 to execute a reaching movement which positions the hand appropriately for grasping. The rest of the model presents the core elements for the understanding of the mirror system. Mirror neurons do not fire when the monkey sees the hand movement or the object in isolation – it is the sight of the hand moving appropriately to grasp or otherwise manipulate a seen (or recently seen) object that is required for the mirror neurons attuned to the given action to fire. This requires schemas for the recognition of both the shape of the hand and analysis of its motion (ascribed in the model to temporal area STSa), and for analysis of the relation of these hand parameters to the location and affordance of the object (assumed to be parietal area PF in what follows).

In the MNS model, the hand state was defined as a vector whose components represented the movement of the wrist relative to the location of the object and of the hand shape relative to the affordances of the object. Oztop and Arbib (2002) showed that an artificial neural network corresponding to PF and F5mirror could be trained to recognize the grasp type from the *hand state trajectory*, with correct classification often being achieved well before the hand reached the object. The modeling assumed that the neural equivalent of a grasp being in the monkey's repertoire is that there is a pattern of activity in the F5 canonical neurons that commands that grasp. During training, the output of the F5 canonical neurons, acting as a code for the grasp being executed by the monkey at that time, was used as the training signal for the F5 mirror neurons to enable them to learn which hand-object trajectories corresponded to the canonically encoded grasps. Moreover, the input to the F5 mirror neurons encodes the trajectory of parts of the hand *relative to the object* rather than the visual appearance of the hand in the visual field. This training prepares the F5 mirror neurons to respond to hand-object relational trajectories even when the hand is of the "other" rather than the "self". However, the model only accepts input related to one hand and one object at a time, and so says nothing about the "binding" of the action to the agent of that action.

With this, we can look in more detail at how G&L characterize the *grasp schema* in terms of various parameters: The *role* parameters are agent, object, object location, and the action itself; the *phase* parameters are initial condition, starting phase, central phase, purpose condition, ending phase, and final state; and finally there is the *manner* parameter. Here *Agent* is an individual, *Object* is a physical entity with the parameters size, shape, mass, degree of fragility, and so on; *Initial condition* is Object Location within peripersonal space; *Starting phase* comprises reaching toward object location and opening effector; *Central phase* comprises closing the effector with force appropriate to fragility and mass; the *Purpose condition* is that the effector encloses the object with a given

manner (a grip determined by parameter values and situation) and the *Final state* is Agent in-control-of object.

However, I view it as mistaken to view Agent and Object as being part of the grasp schema. Rather, the grasp schema is involved in control of the hand movement to exploit an object's affordance (recall FARS) or to recognize such a relationship (recall MNS). This is completely separate from the identification of the agent to whom the hand belongs or the identity or category of the object being grasped (Arbib and Mundhenk, 2005). STS and other temporal regions may recognize, e.g., the face of an agent, but they must be linked to mirror systems in F5 to bind the agent to the observed action. Similarly, the inferotemporal cortex (IT) is required to recognize or classify an object, in distinction to AIP's analysis of its affordances. F5 only knows about grasping an affordance. It does not know what the object is which affords it, nor does it know why the object is being grasped.

Similarly, Rizzolatti and Arbib (1998) view the firing of F5 mirror neurons as part of the code for the cognitive form *GraspA* (Agent, Object), where *GraspA* denotes the specific kind of grasp applied to the Object by the Agent. Again, this is an "action description". If attention is focused on the agent's hand, then the appropriate case structure would be *GraspA* (hand, object) as a special case of *GraspA* (Instrument, Object). Thus, the same act can be perceived in different ways: "Who" grasps vs. "With what" the grasp is made. It is worth noting that the monkey's mirror neurons did not fire when the monkey initially observed the experimenter grasping a raisin with pliers rather than his hand but did come to fire after repeated observation. We thus see the ability to learn new constraints on a "slot" – in this case the observed generalization of the Instrument role from hands alone to include pliers.

The full neural representation of the Cognitive Form *GraspA* (Agent, Object) requires not only the regions AIP, STS, 7a, 7b and F5 mirror included in the MNS model as well as inferotemporal cortex (IT) to hold the identity of the object (as seen in the FARS model) but also regions of, for example, the superior temporal sulcus (STS) not included in MNS which hold the identity of the agent. Fig. 1 schematizes the structures involved in the macaque brain but there are no "Linguistic Forms" in the monkey's brain which allow the monkey to communicate these situations to another monkey other than as a side-effect of performing the action itself.

In the case analysis of Fillmore (1966), the sentence "John hit Mary with his hand" is viewed as the "surface structure" for a case structure hit (John, Mary, John's hand), which is an instance of the case frame hit (agent, recipient, instrument), which makes explicit the roles of "John", "Mary" and "John's hand". However, being able to grasp a raisin is different from being able to say "I am grasping a raisin", and (Rizzolatti and Arbib, 1998) are clear that the neural mechanisms that underlie the doing and the saying are different. However, the case structure lets us see a commonality in the underlying representations, thus helping us understand how a mirror system for grasping might provide an evolutionary core for the development of brain mechanisms that support language. A caveat, however. *A* given semantic representation may be read out in many different ways, even within a single language, so that we must carefully distinguish cognitive and syntactic structure, with the former relatively language-independent (Arbib and Lee, 2007).

EMSH hypothesizes that (a) recognition of manual actions, (b) imitation, and (c) the ability to acquire and use language rest on a nested, evolutionary progression of brain mechanisms. I take seriously our ability to produce myriad novel sentences, seeing the openness of language as both grounding for and emerging from the ability to translate between cognitive structures and verbal structures within the framework of social communication. When (Iacoboni, 2004) urges us "... to consider carefully the incontrovertibly motor elements that are at work in conversation [whose processing] ... require[s] a fast functional architecture not dissimilar to the one needed in motor control." my concern is that most species have "a fast functional architecture [for] motor control," but only humans have language, so that his statement ignores the very differences whose bridging across evolutionary time one needs to explain (Roy and Arbib, 2005).

6. Functional clusters and embodied simulation

G&L explore three parallel parietal-premotor cortical networks, F4→VIP; F5ab→AIP; and F5c→PF which are also important in the FARS and MNS models. They view each as a *functional cluster* – "a cortical network that functions as a unit with respect to relevant neural computations". We would just add that (i) hand movements must be coordinated with arm actions so that a "cluster" is a set of brain circuits chosen by the analyst, rather than being separated from other brain regions in any strong sense; and (ii) inferotemporal cortex and prefrontal cortex play important roles.

6.1. The F4-VIP cluster

Premotor area F4 (a sector of area 6 in the macaque monkey brain) contains neurons that integrate motor, visual, and somatosensory modalities for the purpose of controlling actions in space and perceiving peripersonal space, the area of space reachable by head and limbs (Fogassi et al., 1996). F4 neurons are part of a parietal-premotor circuit which serves not only to control action, but also to construct an integrated representation of actions together with the locations of objects toward which actions are directed. This was modeled as the reach controller in the MNS model. For this cluster, the properties of the object are far less important than its spatial position. Damage to this cluster will result in the inability to be consciously aware of, and interact with, objects within the contralateral peripersonal space (Rizzolatti et al., 2000).

6.2. The F5ab-AIP cluster

F5ab is the region of F5 containing the canonical neurons. This was modeled as the grasp controller in the MNS model. G&L claim that the sight of an object at a given location, or the sound it produces, automatically triggers a "plan" for a specific action directed toward that location in the form of firing of F5 canonical neurons and that the "plan" to act is a simulated potential action.

6.3. The F5c-PF cluster

F5c is the region of F5 containing the mirror neurons. Roughly, 30% of the mirror neurons for grasping recorded in the macaque are "strictly congruent." They fire when the action seen is exactly the same as the action performed. The others fire for a wider range of observation than execution, e.g., one might fire when the monkey *executes* a pincer grip or *observes* any type of grasping. The core of the MNS model was to show how the recognition of grasps could be learned by this cluster. G&L assert that when the subject (a monkey) observes another individual (monkey or human) doing an action, the subject is automatically simulating the same action. However, G&L have already claimed that firing of F5 canonical neurons is a simulated potential action. They do not discuss the differential roles of these 2 distinct representations of an action. Moreover, the existence of mirror neurons which are not strictly congruent suggests that the latter representation may lack conceptual specificity.

G&L recall a series of experiments (Umiltà et al., 2001) in which F5 mirror neurons were tested in two conditions: (1) a condition in which the monkey could see the entire action (e.g., a grasping-action with the hand), and (2) a condition in which the same action

was presented, but its final part was hidden. In the hidden condition, the monkey was shown the target object before an occluder was placed in front of it. More than half of the recorded neurons responded in the hidden condition. G&L assert that "These data indicate that, like humans, monkeys can also infer the goal of an action, even when the visual information about it is incomplete" but this is misleading. The neurons will not fire in case (2) unless the monkey saw the target object before the occluder was positioned. The monkey does *not* infer the goal of the object – he has a *working memory of the object*. This is modeled in MNS2 (Bonaiuto et al., 2007), an extension of the MNS model which addresses not only the (Umiltà et al., 2001) data but also the data (Kohler et al., 2002) on audiovisual mirror neurons. In any case, we have already seen that recognition of the object must be encoded elsewhere.

Whatever their merits as simulations, I doubt that the mere activity of F5 canonical and mirror neurons alone suffices to provide "conceptual understanding" of an action. Consider a pattern recognition device that can be trained to classify pixel patterns from its camera into those which resemble a line drawing of a circle and those which do not (with the degree of resemblance cut off at some arbitrary threshold). It does not *understand* circles. However, to the extent that this recognition could be linked to circuitry for drawing a circle, or for forming associations like "the outline of the sun" or "an orthogonal cut through a cone" as yielding an appropriate stimulus, to that extent can one say that the *system* of which the pattern recognizer is part does exhibit some modicum of understanding. Understanding is thus not a binary concept but rather a matter of degree. Some things may be encoded appropriately yet not understood at all, others may be understood in great richness because their neural encoding is linked to many other behaviors and perceptions.

I agree with G&L that cognitive representations should not be considered in isolation from the motor system; and that macaque neurophysiology has inspired the search for similar systems in the human brain (at the level of imaging of brain regions rather than measurements from single cells). In particular, they cite data that show that some of the same parts of the brain used in seeing are used in visual imagination and some of the same parts of the brain used in action are used in motor imagination. Aziz-Zadeh and Damasio (2008) offer a current assessment of the implications of neuroimaging for theories of embodied cognition. They summarize studies showing that reading words or phrases associated with foot, hand, or mouth actions (e.g., kick, pick, lick) activates (pre)-motor areas adjacent to or overlapping with areas activated by making actions with the foot, hand, or mouth (Aziz-Zadeh et al., 2006; Hauk et al., 2004; Tettamanti et al., 2005). Interestingly (their Fig. 1), the sites of peak activation vary from study to study, posing questions for further research.

All human beings can imagine worlds that they have not seen before, and imagine doing things that they have not done before. Schema theory (ST) provides part of the answer. In actual perception of a scene, we assemble parameterized schemas to match aspects of the sensory input, where the input attended to, and the schemas selected, will depend both on our recent history and the task at hand (perception is action-oriented). Imagination, then, can be seen as the formation of a schema assemblage where the choice of schemas and parameters is freed from the constraints of current sensory input, but may still be affected by the recent history of the assemblage and by our current concerns. Over time, we may extend our stock of schemas beyond those of our own embodied experience – by forming a "mental picture" on hearing a story, by adapting schemas to match pictures, or by building novel abstractions which exploit the combinatorial properties of the symbols we have learned to employ. Some of the novel schemas thus formed, like that for flying, can be seen as extensions of our embodiment; others seem to rest on social context, some linking to our embodied relations to others, and some, as in mathematics or abstract symbolic rules (Arbib, 1990), linked to embodiment only by the fact that we, the cognizers, have a history as embodied beings.

7. Basic-level categories

In a hierarchy like furniture/chair/rocking chair or vehicle/car/sports car, the categories in the middle—chair and car—are what (Rosch, 1978) called "basic-level" categories. Words for basic-level categories tend to be learned earlier, to be shorter, to be more frequent, to be remembered more easily, and so on. G&L add two properties to this list: (i) one can get a mental image of a chair or a car, but not of a piece of furniture in general or a vehicle in general, and (ii) that we have motor programs for interacting with chairs and cars, but not with furniture in general or vehicles in general. They then add that the basic level is the highest level at which this is true. They claim that this implies that categorization is *embodied* – given by our interactions in the world, not just by objective properties of objects. Without the way we sit and the way we form images, the wide range of objects we have called "chairs" do not form a category. However, in the hierarchy *animal/mammal/whale/sperm whale*, the base-level category (whale) does not depend on human existence, let alone embodiment. I chose *whale* here because it is only our layman's access to biological science (not our own experience) that places *whale* in the above hierarchy, rather than in the *fish* hierarchy where unaided visual observation would have us place it. My point here is not to deny the importance of sensorimotor experience in grounding our schemas (recall the seminal work of Piaget, and note his discussion with Beth of the transition via reflective abstraction to mathematical thinking, (Beth and Piaget, 1966)) but rather to emphasize how important it is to understand how much more is involved in categorization.

In saying that categorization is embodied, G&L have included not only the way we sit but also the way we form images. However, the mechanism of using a lens to form an image seems so general a way of transforming the shape and texture of patterns of objects to 2-dimensional arrays that it seems to weaken the idea of embodiment. Thus, although the cones of our retina evolved to support trichromatic vision, it seems unhelpful to say that "green" is an embodied concept. Indeed, people from different cultures divide the color space in different ways, but not because they have different types of bodies – and one can indeed "teach" a given concept of green to any computer equipped with a color camera. However, I do agree that many concepts are indeed defined – like "sit" and "chair" – not by pattern recognition on the basis of vision alone, but rather by the multi-modal integration of multiple senses, actions, and the consequences of actions. Thus, one could design a computer vision system that could *discriminate* with high accuracy pictures of steering wheels from pictures of other objects, but for most of us our *understanding is embodied* – associating the use of the hands to turn the wheel in one direction or another with the correlated vestibular and visual experience of the car turning accordingly.

G&L state their central claim on embodied concepts as follows: "The job done by what have been called "concepts" can be accomplished by schemas characterized by parameters and their values". They then assert that a schema, from a neural perspective, consists of a network of functional clusters, (i) one cluster to characterize each *parameter*, (ii) one cluster for each *parameter value*, or range of values, and (iii) one "controller" cluster, which is active. No mention is made of any of the earlier work on schemas, starting with Head and Holmes (1911) in neurology and Piaget (1971) in psychology, but here I will contrast the G&L notion of schemas

with ST. We have already seen that some schemas may indeed integrate perceptual and motor schemas, but we may in general separate perceptual schemas (e.g., for *apple*) from motor schemas (*e.g.,* for *peeling*) and separate these in turn from more abstract schemas. However, the talk of functional clusters raises considerable problems. We have learned that, for example, F5c-PF is a functional cluster for the mirror system for grasping. But this granularity is quite wrong for more specific concepts related to hand action, like "punch", "tear", "probe", etc. Perhaps, then, G&L would argue that there are dedicated "subclusters" for each of these tasks. The problem with this can be demonstrated by the detailed analysis of schemas for "approach prey" versus "avoid predator" in the frog. Here (Ewert and von Seelen, 1974), we find that the same neurons of the frog (or toad) tectum appear to be involved in recognizing the presence and location of a prey or a predator, but that modulation from the pretectum, activated by recognition of a predator, can inhibit the prey-related effect of tectal output and determine a different motor response to the localized moving stimulus (Cobas and Arbib, 1992). Moreover, there is no separate neural representation of a parameter from the values of the parameter, and it is the activity of the neural coding of these "parameter values" that acts as the overall activity level of the schema. In short, the neural realization of two different schemas – here "approach" and "avoid" – can share neural circuitry (they do not necessarily engage distinct functional clusters), and parameters may or may not be represented separately from their values. More generally, we may note the attempts to re-examine mirror neurons in the context of forward and inverse models (Demiris and Hayes, 2002; Oztop et al., 2006; Wolpert and Kawato, 1998). Each motor schema would correspond to a distinct system (controller + forward model + inverse model) played across G&L's functional clusters, so that the granularity of their analysis seems ill-suited to an account of action concepts, and seems irrelevant to concepts related to classes of objects other than those (e.g., tools) strongly associated with characteristic actions.

In our discussion of the dorsal "how" pathway and ventral "what" pathway, we did see that different "functional clusters" encoded the parametric values that actually guided movement and the "declarable values" relevant to planning and language. However, there is no obligatory neural relation between motor parameters and "parameters we can talk about." In the case of the frog, the latter do not exist. In the case of human hand movements, experience has tuned two separate systems to work together harmoniously, but leaving open the possibility which G&L seek to exclude – namely that language rests on far more than the exploitation of functional clusters present in all primates to subserve sensorimotor integration. However, neither G&L nor ST gives a good feel for how we get from sensorimotor schemas, or perceptual and motor schemas, to abstract concepts like "liberty" or "differential equation". Later, I will evaluate how well G&L use Lakoff's prior work on metaphor to get us started in the right direction.

Let us agree that schemas arise from (1) the nature of our bodies, (2) the nature of our brains, and (3) the nature of our social and physical interactions in the world, and that as a result schemas are not purely internal, nor are they purely representations of external reality. However, this raises three concerns: (i) to think through the difference between a schema as a function of a particular brain and a concept as something which members of a society can share, (ii) to address the fact that besides perceptual and motor schemas there are schemas at varied layers of abstraction, and (iii) that having abstract schemas that contribute to coordinated control programs to praxic behavior is not the same as having linguistic means to express the concepts to which the schemas may be related. Some steps in this direction were presented by Arbib and Hesse (1986) in their presentation of *social schemas*, but these lie outside the scope of the present article.

8. From embodied cognition to the Berkeley "Neural Theory of Language"

G&L believe "that the schemas structuring sensory-motor parameterizations can be used to characterize all concrete concepts. ... What brings together the perceptual and motor properties are, [they] believe, .. [the] characteristics in humans that have been found for *canonical neurons* thus far in the brains of monkeys." The catch with localizing concepts in canonical neurons is that (a) concepts concerning categories of objects are not associated with unequivocal actions (recall our discussion of apples) and (b) canonical neurons do not fire when we observe an action – and this would seem to invalidate them as candidates for concept representation.

G&L review the computational neural models of Feldman and Narayanan (2004) and Narayanan (1997) for motor schemas, which include premotor, motor, and premotor–motor connections. The premotor model functions dynamically to "choreograph" and carry out in proper sequence simple movements encoded in motor cortex. These premotor models are given a uniform structure: (1) initial state, (2) starting phase transition, (3) precentral state, (4) central phase transition (either instantaneous, prolonged, or ongoing), (5) postcentral state, (6) ending phase transition, and (7) final state. At the postcentral state, there are the following options: (a) a check to see if a goal state has been achieved, (b) an option to iterate or continue the main process, (c) an option to stop, and (d) an option to resume. (This is reminiscent of the classic TOTE (Test-Operate–Test-Exit) units of Miller et al. (1960).) Each complex motor program is a combination of structures of this form, either in sequence, in parallel, or embedded one in another. What distinguishes actions from one another is (i) the version of this premotor structure and (ii) bindings to the motor cortex and other sensory areas (for perceptual and somatosensory feedback). These premotor structures are called "executing schemas," or X-schemas and seem to be an instantiation in structured connectionist neural networks of the motor schemas and coordinated control programs of ST. Note that the FARS model (Fagg and Arbib, 1998) provides a model of simple sequencing of motor actions that links to the biology of F5 canonical neurons and AIP, but also involves basal ganglia and various areas of prefrontal cortex. G&L assert that "Narayanan (1997) noted that premotor structures also fit the perceptual structure of the motor actions modelled. In short, he modelled the structures described above for mirror neurons, canonical neurons, and action-location neurons" though in fact no mention of mirror neurons or canonical neurons is made in Narayanan's thesis.

Narayanan's real contribution is not to neurobiology but rather that he offers a theory of conceptual metaphor based on computations in structured connectionist networks. Conceptual metaphors are one of the basic mechanisms of mind (Arbib and Hesse, 1986; Lakoff and Johnson, 1980). Each conceptual metaphor is a mapping across conceptual domains. For example, the conceptual metaphor *love is a journey* maps travelers to lovers, vehicles to relationships, destinations to common life goals, and impediments to relationship difficulties, as shown by English expressions about love like *It's been a long bumpy road* or *The marriage is on the rocks.* Love can also be conceptualized and reasoned about in terms not only of a journey, but also of a partnership, a joining-together, magic, heat, and so on. For Lakoff and Johnson (1980) most of the richness of the concept comes from these metaphors, but note that, contrary to G&L, this "Lakovian" view of metaphor does not restrict the target domain to the sensorimotor realm. In any case, G&L give us no sense of how "marriage" is to be distinguished from being "just a journey". Surely, metaphor rests on the appreciation of difference as much as commonality. My point is not to deny that some apparently abstract concepts may be linked to sensorimotor

concepts; rather it is to deny that human concepts are exhausted by what can be conducted in the sensorimotor linkages of parietal and premotor cortex.

Narayanan (1997) constructed a computational neural model of metaphorical mappings and implemented conceptual metaphors mapping physical actions to economics, processing sentences like *France fell into a recession* which include physical sensory-motor expressions like *fall into*. He then showed that his model could use the mappings to combine source (sensory-motor) and target (economic) inferences to get correct inferences from such sentences. This is an interesting result in connectionist inferencing, but does not justify G&L's claim that "concepts characterized in the sensory-motor system are of the right form to characterize the source domains of conceptual metaphors." What, for example, is the sensorimotor grounding for the concept *France* in the above sentence?

An important feature of Narayanan's X-schemas, lacking in the schemas of ST, is that their structure supports a simple analysis of *aspect*. These indicate the duration of the activity described by a verb, such as *about to_Verb, start to_Verb, be_Verb_ing, and have_Verb_Past Participle*. Specifically, this involves including the before and after states within each X-schema. However, the memory of doing something is not *within* the motor schema though it may invoke it. Moreover, adding these states to a motor schema provides little help in representing the concepts invoked in a sentence like "The second time he threw the ball at the target he succeeded, but the third time he dropped it." Anyway, let's see how G&L exploit these extra states of the X-schemas. Premotor cortex is neurally connected to the motor cortex, but those connections can be inhibited, and the X-schemas of the premotor system can function independently. A sentence like *he is doing something stupid* does not tell what action he is carrying out, but does specify the ongoing aspectual structure, with the inferences that *he has already started doing something stupid* and *he hasn't finished doing something stupid*. G&L state that these inferences are being computed via neural simulation by X-schema structure circuitry in the premotor cortex, with no active connections to the motor cortex. However, since the action is not specified, this requires that there be an X-schema for *doing,* an assumption which seems dubious. It is hard to see what simulating "doing' entails especially since the *doing* need not be physical at all, and there is no empirical support for G&L's claim to have shown that "the premotor cortex is being used to do abstract reasoning, reasoning that is not about any particular sensorimotor activity." Of course, I need to confess that I have no satisfactory theory of verb aspect either. Thus Narayanan's explicit model stands as a challenge to other studies which, like G&L and EMSH, seek to chart the evolution of language within the framework of prior mechanisms for action and the perceptual mechanisms which serve it.

G&L use the term *secondary area* for a brain region not directly connected to sensors or effectors, but which provides structure to information going to effectors or coming from sensors. This is a rather poorly couched definition, since none of cerebral cortex is connected to sensors or effectors. Given the definition of their functional clusters, one might infer that for G&L the secondary areas are premotor and parietal. If so, this would, for example, exclude IT and PFC. In any case, G&L tell us that Lakoff has proposed (no reference is given) a generalization of Narayanan's account of abstract aspectual concepts to define a *cog* as any concept that satisfies the following:

- Cogs (which include abstract aspectual concepts) are neurally simulated in a secondary area, with no active connections to a primary area. Emphasis is placed on premotor cortex when all connections to motor cortex are inhibited.

This is confusing since it now suggests that we do not have, for example, the abstract concept of an action if we are actually executing that action. Indeed, in apparent contradiction of the above formulation, G&L later state that "These concepts are general, and can apply to any special-case concepts when there are active connections to primary areas. When there are such active connections, the general concepts are an inseparable part of the structure of the special case concepts."

- Inferences about cogs are computed via that simulation.

This seems to suggest that inference takes place within premotor cortex, but the case is not proven. How might premotor cortex support an inference like "Today will be yesterday tomorrow"?

- Cogs characterize concepts in the grammar of a natural language.

This seems either trivial or false. If it just means that the grammars of many natural languages mark concepts like aspect, this is trivial. But if it claims that X-schemas are structured according to the way that aspect, say, is marked in the speaker's language, it seems to be false, since aspect is marked in different ways, if marked at all, in different languages. More generally, it seems mistaken to talk of any concept as being "in the grammar of a natural language". I would suggest that my concept of *blue* remains the same whether I am speaking English and place the word "blue" before the noun or speaking French and place "bleu" after the noun.

G&L state that cogs may include all the primitive image-schemas, such as containment, source-path-goal, force dynamics, and orientation schemas (Langacker, 1986, 1991; Talmy, 1985) but no evidence has been given as to which secondary areas, if any, form a cluster for computing these concepts. More to the point, G&L continue to ignore the distinction between "having a concept" and "having a form of words to express that concept." None of containment, source–path–goal, force dynamics, or orientation schemas is a grammatical construct, though each certainly plays a role in the definition of *cognitive grammar* (Langacker, 1986, 1991). Rather, different languages will employ different grammatical forms (e.g., a preposition versus a postposition for containment) to express them. Again, even someone who adheres to "early cognitivism," namely that concepts are symbolic representations by nature, would concede that one can perceive that a scene one is looking at falls under a concept, and can act on the basis of that perception. Thus the linkage of conceptual machinery to secondary areas is irrelevant to the debate over whether or not the concept is embodied.

Yet G&L go on to say "If [such concepts] are all computed in secondary regions, that would explain why there is a limited range of them (there is a relatively small number of such regions), why they are universal (we all have the same basic brain structure) and why they are general (they provide structure to primary regions with specific information)." But brains are plastic. I think G&L place too much weight on shared brain structure and too little on shared experience that shapes these structures. They further assert (p. 471):

> Because neural structures in secondary areas are inseparable in behavior from the primary structures that they are connected to, they characterize generalizations that are inherent in and inseparable from special cases. The "learning" of general cases is not the acquisition of new structures, but rather the inhibition of the connections between secondary and primary areas. The generalizations are inherent in the special cases that are learned first. What is learned is the control of inhibitory connections.

But different subsets of special cases will support different generalizations. And once we see a pattern, in what sense is this a mat-

ter of inhibition save in the trivial sense that all neural computation involves inhibition as well as excitation? No details are given to show how merely inhibiting connections between secondary and primary areas can yield the representation of a concept like "tomorrow".

They predict that in fMRI studies, a metaphorical sentence like *He grasped the idea* should activate the sensory-motor *grasping-related regions* of the brain. Similarly, a metaphorical sentence like *They kicked him out of class* should active the sensory-motor *kicking-related regions* of the brain. However, this conflates the use of a "fresh metaphor" whose meaning we can gather only by invoking the source domain with a "frozen metaphor" whose implications are well-known and have become stored as an enrichment of the target domain. Indeed Aziz-Zadeh et al. (2006) find that when subjects read frozen metaphors related to actions (e.g., "bite the bullet", "grasp the meaning", "kick the bucket"), no congruent somatotopic organization of semantic representations (of mouth, hand and leg areas, respectively) was found. Such data seem to count against G&L's strong view of embodied cognition, while still allowing the more restricted view that some of our concepts are indeed strongly embodied, and that the neural realization of these concepts has a strong effect on the corresponding sensorimotor circuitry, whether or not those concepts (and related linguistic structures) are entirely supported by the same sensorimotor circuitry.

9. Back to the mirror system hypothesis

G&L see it as an argument against the traditional modality-neutral, disembodied account of concepts that in order to have a neural account of such a theory, *action concepts*, like all other concepts, would have to be represented neurally *outside the sensory-motor system altogether*. They see it as a counter-argument that this would require duplication of the neural correlates that neuroscience has found in the sensory-motor system for manner subcases, the agent–object–location structure, the purpose structure, and the phase structure. However:

(i) Pronouncing a word that describes an action is very different from executing the action. The neural control for saying *run* is very different from the neural control for running, and most creatures that can run are unable to produce a symbol that denotes running. EMSH sees portions of the language-ready brain as derived from the general primate *sensory-motor system* but denies that that system in and of itself is adequate to support language.

(ii) Even if there are some economies in linking action parameters to the verbal description of their values, the fact remains that English contains nouns, adjectives, determiners, prepositions– not to mention the constructions which combine them – that have no *immediate* roots in the motor system. Indirectly, one needs ways to describe agents, objects, actions and the relation between them. We have a motor system for controlling eyes, legs, and hands for example, and each language must provide a way to describe events that combine them. But the very diversity of ways in which languages express this (and whether in speech, sign language or some multimodal form of communication) suggests that there is no direct linkage from sensorimotor to linguistic construction. In my view, once protolanguage was established, different peoples developed (and later shared) different strategies for talking about things and actions, and then developed these strategies in diverse ways to talk about more and more of their world. Indeed, some languages, like Vietnamese, lack all inflection, precluding the use of inflectional criteria for identifying grammatical categories; while other languages employ inflection in unusual ways (Kemmerer, 2005). For example, the language of the Makah of the Northwest coast of the United States (http://www.native-languages.org/makah.htm) applies aspect and mood markers not only to words for actions that are translated into English as verbs, but also to words for things and properties.

I suggest that the verbalization of an action is just as indirect as the verbalization of a noun. To understand this, note that EMSH does not claim (as suggested by Barrett et al. (2005)) that the exact same machinery used to act and recognize an action supports the naming of that action and the recognition of that name. Rather, it suggests that the evolution of the human brain saw major variations on the theme of the mirror system and the brain regions with which it interacts (Arbib, 2006). EMSH traces a path from praxic imitation to pantomime in the service of communication to the development of conventionalized protosign and thence speech. Note that pantomime does not privilege action – one can communicate about an object not only by pantomiming a typical action in which it is involved but also by tracing its shape, for example. On the other hand, there are concepts like "green" for which no immediate embodiment or pantomime is available – and, presumably, it was the utility of communicating about such concepts that made it an adaptive part of our evolutionary history to move beyond pantomime to the conventionalized gestures of protosign and the even more conventionalized gestures of protospeech, with the human facility for pointing playing a crucial role in drawing attention to exemplars to be used in learning to recognize the concept associated with a given protoword.

Fig. 2 (Arbib, 2006) makes explicit that, developing EMSH, we postulate that a mirror system for phonological expression ("words") evolved atop the mirror system for grasping to serve communication integrating hand, face, and voice. Additionally, following ST, we postulate that the concepts for diverse actions, objects, attributes, and abstractions are represented by a network of schemas stored in LTM (long-term memory), with our current "conceptual content" formed as an assemblage of schema instances in working memory (WM) in a system that includes inferotemporal cortex (IT) and prefrontal cortex (PFC), where the latter, at least, appears not to be a secondary area in the sense of G&L. Analogously, the Mirror for Words contains a network of word forms in LTM and keeps track of the current utterance in its own working memory (cf. Baddeley's phonological buffer, Baddeley, 2003). The perhaps surprising aspect of Fig. 2 is that the arrow linking the "Mirror for Actions" to the "Mirror for Words" expresses, it is hypothesized, an evolutionary relationship, not a flow of data. Rather than a direct linkage of the dorsal representation of an action to the dorsal representation of the phonological form, we have two relationships between the dorsal pathway for the Mirror for Actions and the schema networks and assemblages of the ventral pathway and prefrontal cortex (PFC). The rightmost path shown in Fig. 2 corresponds to the paths in the FARS model whereby IT and PFC can affect the pattern of dorsal control of action. The path just to the left of this shows that the dorsal representation of actions can only be linked to verbs via schemas. In accordance with this view, Hickok and Poeppel (2004) offer data on cortical stages of speech perception that distinguish a dorsal stream mapping sound onto articulatory-based representations from a ventral stream mapping sound onto meaning.

Moreover, the MNS model suggests that the correct firing of mirror neurons is a learning process – it is the adaptation of connections from STS and PF that adapts a mirror neuron to fire when the agent observes another perform one of a particular class of actions (Bonaiuto et al., 2007; Oztop and Arbib, 2002). According to this model (which is consistent with available data) the linkage

of a visual (or other perceptual) concept with a motor concept occurs not so much because of direct connections within the brain but because the perceptual and motor systems can correlate the activities that occur in each of them while the organism is interacting with the world. But when it comes to non-motoric concepts like "green" or "avocado" or "advocate" it is the brain's ability to correlate neural patterns based on perceptual input with the neural patterns for specific words and phrases that plays the crucial role in the individual's concept formation, and embodiment plays at most a secondary role, if any at all.

10. Towards abstraction

Overall, G&L seem to claim that concepts are either premotor or image-schemas or reducible to such via metaphor. No adequate account has been given of how such processes could support abstract concepts such as causation, identity and love. I don't think they can. Moreover, it is striking that they discuss virtually none of the brain beyond VIP/AIP/PF and F4/F5, whereas most scholars would ascribe vital roles to temporal and prefrontal cortices in concept formation. To be candid, I can offer no easy route to embedding causation, identity and love in EMSH. What I shall try to do, though, is to suggest how one might build a path from embodiment to abstraction. In this view, the systems charted by G&L remain important, but are only part of larger systems that supports the full range of human cognition and language.

Bickerton (1995) posed an important challenge to too direct a linkage of sentences to sensorimotor representations. He notes (p. 22) that a sentence like "The cat sat on the mat" is far more abstract than the image of a particular cat sitting on a particular mat. Moreover, an image does not bring in the sense of time distinguishing "The cat sat on the mat" from "The cat is sitting on the mat" or "The cat will sit on the mat". Again, an image would not distinguish "The cat is sitting on the mat" from "The mat is underneath the cat". All this is true, and we must reflect these distinctions in characterizing language. For example, we might relate the focus of a sentence (where prosody plays a crucial role not obvious in the written words) to the focus of attention in vision (Itti and Arbib, 2006). However, Bickerton creates a false dichotomy when he asserts that "it is not true that we build a picture of the world and dress it out in language. Rather, language builds us the picture of the world that we use in thinking and communicating." The idea that language builds our picture of the world – rather than contributing to its richness – is misguided for it ignores the role of visual experience and then of *episodic memory* (linking episodes in temporal and other relationships) and *expectations* in building the rich perceptions and cognitions (Cognitive Form) of which sentences (Phonological Form) are just a précis. There is no claim that the relationship is one-to-one. Bickerton's approach leaves little room for understanding how the ability to *mean* that a cat is on the mat could be acquired in the first place. The language of the brain or schema network is vastly richer than a linear sequence of words. This does not deny that language can express what pictures cannot or vice versa. Note that perception is *not* invertible: even if I see an actual cat on an actual mat, I am unlikely to recall more than a few details. And what one sees is knowledge-based: e.g., a familiar cat vs. a generic cat, or recognizing a specific subspecies. There is an intimate relation between naming and correct categorization.

Bickerton (1995, pp. 22–24) argues that one cannot picture "My trust in you has been shattered forever by your unfaithfulness." because no picture could convey the uniquely hurtful sense of betrayal the act of infidelity provokes if you did not know what trust was, or what unfaithfulness was, or what it meant for trust to be shattered. "In the case of trust or unfaithfulness, there can be nothing beneath the linguistic concept except other linguistic representations, because abstract nouns have no perceptual attributes to be attached to them and therefore no possible representation outside those areas of the brain devoted to language." This is wrong on three levels:

(i) The words themselves (i.e., the sequences of letters on the page or spoken phonemes) do not convey "the uniquely hurtful sense of betrayal the act of infidelity provokes". They only convey such feelings if they "hook into" an appropriate body of experience and association, which not all people will share – the word is the "tip of the schema iceberg". Words must link into the network which itself links to non-verbal experience, both perceptual and behavioral (cf. the discussion of a person's knowledge as a "schema encyclopedia" in Arbib (1985, p. 43)).

(ii) Given this, an image (whether static like a picture, or extended in time like a video clip) may tap a similar network of experience, such as seeing one person turning away with an expression of disillusionment and despair from the sight of another engaged in lovemaking. The words and the images have complementary strengths – the words make explicit the key relationships, the image provides a host of details that could be only supplied in language (if indeed they were deemed relevant) by the piling on of more and more sentences.

(iii) If one recalls a beautiful sunset, then it may be that "The sunset where we saw the green flash at Del Mar" will *index* the scene in my own thoughts or for communication with others, but the words alone do not recapture the beauty of the scene by forming an image of the setting and the colors of the sky.

As an exercise, let me try to link the sentence "My trust in you has been shattered forever by your unfaithfulness" back to the schema network anchored in my action and perception. I look at the definitions of the words and see how they are – eventually – rooted in behavior, noting the necessary role of metaphor in the use of "shattered", and in the use of "your" to indicate both possession of an object and possession of a disposition:

My trust in you expresses the objectification of the behavioral schema *Trust (I, You)*, where *Trust(A,B)* means "For all *C*, *B* tells *A* that *C* is the case $\Rightarrow$ *A* acts on the assumption that *C* is true". (I do *not* argue that my mental states need exploit representations expressing such a formalism. Rather, the above formula is a short-hand for a whole range of behaviors and expectations that constitute the mental state of "trusting".)

B is faithful to A is defined socially by a set of behaviors *prescribed* and *proscribed* for *B* by the nature of his/her relationship to *A*. Infidelity is then detected by, perhaps, repeated failure in a prescribed behavior, or possibly even one example of proscribed behavior.

That an object is *broken* is, in the grounding case, testable either perceptually – the recognizable structure has been disrupted – or behaviorally – the object does not behave in the expected way. *Repairing* is acting upon an object in such a way as to make it look or perform as it is expected to. An object is *shattered* if it is broken into many pieces – implying that repairing the damage (making the object functional again) will be difficult or impossible.

Shattered forever then asserts that repair is impossible – there is no set of operations such that at any future time the object will function again, introducing the element of time and a hypothetical, involving the *semantic extension of schemas from the here and now of action and perception*. But note too that *planning* and *expectations* are implicit in behavior, and relate to the notion of an *internal model of the world*. Moreover, our notions of future time rest on

extrapolation from our experience of past times in relation to the expectations we held at even earlier times.

Having said all this, note the many "inventions" required to go from simple wants and actions to a language + thought system rich enough to express the above sentence. But note, too, that the formal aspects sketched above do not begin to exhaust the meaning of the above sentence, and this can only be done by consideration of the embodied self. To say my "trust is shattered" also implies a state of emotional devastation that needs empathy of another human to understand.

This account is little more than a caricature but serves to reinforce the view that the use of language is rooted in our experience of action within the world, enriched by our ability to recall past events or imagine future ones and expanded by our cultural history as reflected in our own personal experience. The ability to understand "My trust in you has been shattered forever by your unfaithfulness." is not the expression of some standalone linguistic faculty serving representation rather than communication, but expresses the fruits of our individual cognitive and linguistic development within a community that has built a rich set of linguistic devices through an expanding spiral increasing the range of language and cognition through their mutual interaction.

11. Conclusions

Putting all this together, I suggest that the combined study of G&L and EMSH lets us reach the following conclusions:

1. Language does indeed make use in large part of brain structures akin to those used to support perception and action in praxis, with mirror systems providing the core support for parity – matching meaning between "speaker" and "hearer". However, one must carefully distinguish concepts from the linguistic forms which express them. The ability to generate and recognize the phonology of language evolved in part from the canonical and mirror systems for grasping, but the actions of producing words and sentences gain their meaning by their linkage to a distinct structure of schemas that has evolved (both ontogenetically and phylogenetically – the former in the spirit if not the letter of (Piaget, 1954, 1971)) from basic sensorimotor schemas.
2. The human language-ready brain rests on evolutionary innovations that extend far back in the primate line, including possessing a mirror system for dexterous manipulation, and then to complex imitation, pantomime, protosign and the extension of protosign and protospeech. However, language – as distinct from protolanguage – is unique to humans – as distinct from protohumans – and reflects a rich history of human inventions of tools for increasingly subtle communication.
3. There is no such thing as a "language module" if by that is meant an encapsulated module in the sense of Fodor (1983). However, the human brain has structures that distinguish it from the brains of other species and without which the development of language would not have occurred.
4. Grammar provides the means to express not only conceptual schemas but also novel schema assemblages by appropriate combinations of phonological schemas. However, since animals have concepts but no language, and since different languages express concepts in very different ways, the hierarchy of grammatical structure is different from that for conceptual structure. Indeed, grammar lies not so much in the connection between specific concepts so much as in the power to create novel sentences by linking words in hitherto unimagined ways. This seems qualitatively different from, for example, our ability to grasp novel objects.
5. The semantics of grammar involves far more than the sensory motor system, proceeding from relatively direct linkage of items of the lexicon to sensorimotor experience for everyday objects and human actions to more elaborate form-function mappings based on constructions, and to patterns of increasing abstraction based in part, but by no means exclusively, on metaphor.
6. Semantics and grammar have their roots in specific sensorimotor experience but have developed (both historically and ontogenetically) through layer upon layer of abstraction to handle concepts which are not embodied save through their history. (Consider, by analogy, the way in which one may use a word for years before learning its etymology.)
7. Syntax and semantics (which we view as integrated via the form-meaning pairs of constructions) provide a symbolic overlay for our embodied experiences as well as counterfactual events, generalizations and abstractions, and provide a whole range of speech acts which extend the possible range of social interactions.

References

Aboitiz, F., 1995. Working memory networks and the origin of language areas in the human brain. Medical Hypotheses 44 (6), 504–506.

Aboitiz, F., Garcia, R.R., Bosman, C., Brunetti, E., 2006. Cortical memory mechanisms and language origins. Brain Language 98 (1), 40–56.

Aboitiz, F., Garcia, V.R., 1997. The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. Brain Research and Brain Research Review 25 (3), 381–396.

Arbib, M.A., 1981. Perceptual structures and distributed motor control. In: Brooks, V.B. (Ed.), Handbook of Physiology – The Nervous System II. Motor Control. American Physiological Society, pp. 1449–1480.

Arbib, M.A., 1987. Modularity and interaction of brain regions underlying visuomotor coordination. In: Garfield, J.L. (Ed.), Modularity in Knowledge Representation and Natural Language Understanding. The MIT Press, pp. 333–363.

Arbib, M.A., 1990. A Piagetian perspective on mathematical construction. Synthese 84, 43–58.

Arbib, M.A., 2002. The mirror system, imitation, and the evolution of language. In: Dautenhahn, K., Nehaniv, C.L. (Eds.), Imitation in Animals and Artifacts. Complex Adaptive Systems. MIT Press, pp. 229–280.

Arbib, M.A., 2005a. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics (with commentaries and author's response). Behaviour in Brain Science 28, 105–167.

Arbib, M.A., 2005b. Interweaving protosign and protospeech: further developments beyond the mirror. Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems 6, 145–171.

Arbib, M.A., 2006. Aphasia, apraxia and the evolution of the language-ready brain. Aphasiology 20, 1–30.

Arbib, M.A. 2007. How new languages emerge (Review of D. Lightfoot, 2006, How New Languages Emerge, Cambridge University Press). Linguist List 18-432, Thu Feb 08 2007, <http://linguistlist.org/issues/17/17-1250.html>.

Arbib, M.A., 2008. Holophrasis and the Protolanguage Spectrum. Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems 9, 151–165.

Arbib, M.A., Bota, M., 2003. Language evolution: neural homologies and neuroinformatics. Neural Networks 16, 1237–1260.

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

Arbib, M.A., Hill, J.C., 1988. Language Acquisition: Schemas Replace Universal Grammar. In: Hawkins, J.A. (Ed.), Explaining Language Universals. Basil Blackwell, pp. 56–72.

Arbib, M.A., Lee, J., 2007. Vision and action in the language-ready brain: from mirror neurons to Semrep. In: Mele, F. (Ed.), BVAI 2007 (Brain Vision & Artificial Intelligence, 2007), LNCS, vol. 4729. Springer-Verlag, pp. 104–123.

Arbib, M.A., Liebal, K., Pika, S., submitted for publication. Primate vocalization, ape gesture, and human language: an evolutionary framework.

Arbib, M.A., Mundhenk, T.N., 2005. Schizophrenia and the mirror system: an essay. Neuropsychologia 43 (2), 268–280.

Arbib, M.A., Rizzolatti, G., 1997. Neural expectations: a possible evolutionary path from manual skills to language. Communication and Cognition 29, 393–424.

Aziz-Zadeh, L., Damasio, A.R., this issue. Embodied semantics for actions: findings from functional brain imaging. Journal of Physiology, Paris.

Aziz-Zadeh, L., Wilson, S.M., Rizzolatti, G., Iacoboni, M., 2006. Congruent embodied representations for visually presented actions and linguistic phrases describing actions. Current Biology 16 (18), 1818–1823.

Baddeley, A., 2003. Working memory: looking back and looking forward. Nature Reviews Neuroscience 4, 829–839.

Baker, M., 2001. The Atoms of Language: The Mind's Hidden Rules of Grammar. Basic Books.

Baldwin, J.M., 1896. A new factor in evolution. American Naturalist 30, 441–451. 536–533.

Barrett, A.M., Foundas, A.L., Heilman, K.M., 2005. Speech and gesture are mediated by independent systems. Behavioral and Brain Sciences 28, 125–126.

Beth, E.W., Piaget, J., 1966. Mathematical Epistemology and Psychology (translated from the French by W. Mays).

Bickerton, D., 1995. Language and Human Behavior. University of Washington Press.

Bonaiuto, J., Rosta, E., Arbib, M., 2007. Extending the mirror neuron system model, I: audible actions and invisible grasps. Biological Cybernetics 96, 9–38.

Bridgeman, B., 2005. Action planning supplements mirror systems in language evolution. Behavioral and Brain Sciences 28, 129–130.

Byrne, R.W., 2003. Imitation as behavior parsing. Philosophical Transactions of the Royal Society of London (B) 358, 529–536.

Byrne, R.W., Byrne, J.M.E., 1993. Complex leaf-gathering skills of mountain gorillas (*Gorilla* g. *beringei*): variability and standardization. American Journal of Primatology 31, 241–261.

Chomsky, N., Lasnik, H., 1993. The theory of principles and parameters. In: Jacobs, J., von Stechow, A., Sternefeld, W., Vennemann, T. (Eds.), Syntax: An International Handbook of Contemporary Research, De Gruyter, Berlin, pp. 506–556.

Cobas, A., Arbib, M.A., 1992. Prey-catching and predator-avoidance in frog and toad: defining the schemas. Journal of Theoretical Biology 157, 271–304.

Coudé, G., Ferrari, P.F., Roda, F., Maranesi, M., Veroni, V., Monti, F., Rizzolatti, G., Fogassi, L., 2007. Neuronal responses during vocalization in the ventral premotor cortex of macaque monkeys. Society for Neuroscience Annual Meeting. San Diego, California, Abstract 636.3.

Craik, K.J.W., 1943. The Nature of Explanation. Cambridge University Press.

Deacon, T.W., 1997. The Symbolic Species: The Co-evolution of Language and the Brain. W.W. Norton.

Deacon, T.W., 2003. Multilevel selection in a complex adaptive system: the problem of language origins. In: Weber, B., Depew, D. (Eds.), Evolution and Learning: The Baldwin Effect Reconsidered. The MIT Press, pp. 81–106.

Deacon, T.W., 2007. The Evolution of Language Systems in the Human Brain. In: Kaas, J.H., Preuss, T.M. (Eds.), Evolution of Nervous Systems, A Comprehensive Reference. Primates, vol. 4. Elsevier, pp. 529–547.

Demiris, Y., Hayes, G., 2002. Imitation as a dual-route process featuring predictive and learning components: a biologically-plausible computational model. In: Dautenhahn, K., Nehaniv, C. (Eds.), Imitation in Animals and Artifacts. MIT Press.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., 1992. Understanding motor events: a neurophysiological study. Experimental in Brain Research 91 (1), 176–180.

Dominey, P.F., Hoen, M., 2006. Structure mapping and semantic integration in a construction-based neurolinguistic model of sentence processing. Cortex 42 (4), 476–479.

Ewert, J.-P., von Seelen, W., 1974. Neurobiologie und System-Theorie eines visuellen Muster-Erkennungsmechanismus bei Kroten. Kybernetik 14, 167–183.

Fagg, A.H., Arbib, M.A., 1998. Modeling parietal–premotor interactions in primate control of grasping. Neural Networks 11 (7–8), 1277–1303.

Feldman, J., Narayanan, S., 2004. Embodied meaning in a neural theory of language. Brain Language 89 (2), 385–392.

Ferrari, P.F., Gallese, V., Rizzolatti, G., Fogassi, L., 2003. Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. European Journal of Neuroscience 17 (8), 1703–1714.

Ferrari, P.F., Maiolini, C., Addessi, E., Fogassi, L., Visalberghi, E., 2005. The observation and hearing of eating actions activates motor programs related to eating in macaque monkeys. Behaviour in Brain Research 161 (1), 95–101.

Fillmore, C.J., 1966. The case for case. In: Bach, E., Harms, R.T. (Eds.), Universals in Linguistic Theory. Holt, Rinehart and Winston, pp. 1–88.

Fodor, J., 1983. The Modularity of Mind. A Bradford Book/The MIT Press.

Fogassi, L., Ferrari, P.F., 2004. Mirror neurons, gestures and language evolution. Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems 5, 345–363.

Fogassi, L., Gallese, V., Fadiga, L., Luppino, G., Matelli, M., Rizzolatti, G., 1996. Coding of peripersonal space in inferior premotor cortex (area F4). Journal of Neurophysiology 76, 141–157.

Gallese, V., 2003. The manifold nature of interpersonal relations: The quest for a common mechanism. Philosophical Transactions of the Royal Society of London B 358.

Gallese, V., Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. Trends in Cognitive Science 2, 493–501.

Gallese, V., Lakoff, G., 2005. The brain's concepts: the role of the sensory-motor system in reason and language. Cognitive Neuropsychology 22, 455–479.

Goldberg, A.E., 2003. Constructions: a new theoretical approach to language. Trends in Cognitive Science 7 (5), 219–224.

Goldman, A., 2006. Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading. Oxford University Press.

Goodale, M.A., Milner, A.D., 1992. Separate visual pathways for perception and action. Trends in Neuroscience 15, 20–25.

Gregory, R.L., 1969. On how so little information controls so much behavior. In: Waddington, C.H. (Ed.), Towards a Theoretical Biology. Sketches, vol. 2. Edinburgh University Press.

Grillner, S., Wallen, P., 2002. Cellular bases of a vertebrate locomotor system-steering, intersegmental and segmental co-ordination and sensory control. Brain Research Reviews 40, 92–106.

Hauk, O., Johnsrude, I., Pulvermuller, F., 2004. Somatotopic representation of action words in human motor and premotor cortex. Neuron 41 (2), 301–307.

Head, H., Holmes, G., 1911. Sensory disturbances from cerebral lesions. Brain 34, 102–254.

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

Hill, J.C., 1983. A computational model of language acquisition in the two-year-old. Cognition and Brain Theory 6, 287–317.

Iacoboni, M., 2004. Understanding others: imitation, language, empathy. In: Hurley, S., Chater, N. (Eds.), Perspectives on Imitation: From Cognitive Neuroscience to Social Science: Mechanisms of Imitation and Imitation in Animals, vol. 1. MIT Press.

Itti, L., Arbib, M.A., 2006. Attention and the minimal subscene. In: Arbib, M.A. (Ed.), Action to Language via the Mirror Neuron System. Cambridge University Press, pp. 289–346.

Jacob, P., Jeannerod, M., 2005. The motor theory of social cognition: a critique. Trends in Cognitive Sciences 9, 21–25.

Jeannerod, M., Decety, J., Michel, F., 1994. Impairment of grasping movements following a bilateral posterior parietal lesion. Neuropsychologia 32 (4), 369–380.

Jürgens, U., 1979. Neural control of vocalizations in nonhuman primates. In: Steklis, H.D., Raleigh, M.J. (Eds.), Neurobiology of Social Communication in Primates. Academic Press, pp. 11–44.

Jürgens, U., 2002. Neural pathways underlying vocal control. Neuroscience and Biobehavioral Reviews 26 (2), 235–258.

Kemmerer, D., 2005. Against innate grammatical categories. Behavioral Brain Sciences. <http://www.bbsonline.org/Preprints/Arbib-05012002/Supplemental/>.

Kirby, S., 2000. Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In: Knight, C., Studdert-Kennedy, M., Hurford, J.R. (Eds.), The Evolutionary Emergence of Language. Cambridge University Press.

Kohler, E., Keysers, C., Umiltà, M.A., Fogassi, L., Gallese, V., Rizzolatti, G., 2002. Hearing sounds, understanding actions: action representation in mirror neurons. Science 297, 846–848.

Lakoff, G., Johnson, M., 1980. Metaphors We Live By. University Of Chicago Press.

Lakoff, G., Johnson, M., 1999. Philosophy in the Flesh. Basic Books.

Langacker, R.W., 1986. Foundations of Cognitive Grammar, vol. 1. Stanford University Press.

Langacker, R.W., 1991. Foundations of Cognitive Grammar, vol. 2. Stanford University Press.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. Psychological Review 74, 431–461.

Lightfoot, D., 2006. How New Languages Emerge. Cambridge University Press.

MacKay, D.M., 1966. Cerebral organization and the conscious control of action. In: Eccles, J.C. (Ed.), Brain and Conscious Experience. Springer-Verlag, pp. 422–440.

MacNeilage, P.F., Davis, B.L., 2005. The frame/content theory of evolution of speech: comparison with a gestural origins theory. Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems 6, 173–199.

Mahon, B.Z., Caramazza, A., 2005. The orchestration of the sensory-motor systems: clues from neuropsychology. Cognitive Neuropsychology 22 (3/4), 480–494.

Miller, G.A., Galanter, E., Pribram, K.H., 1960. Plans and the Structure of Behavior. Holt, Rinehart & Winston.

Minsky, M.L., 1965. Matter, Mind and Models. In: Information Processing 1965. Proceedings of IFIP Congress 65, vol. 1. Spartan Books, pp. 45–59.

Mountcastle, V.B., 1978. An organizing principle for cerebral function: the unit module and the distributed system. In: Edelman, G.M., Mountcastle, V.B. (Eds.), The Mindful Brain. The MIT Press, pp. 7–50.

Myowa-Yamakoshi, M., Matsuzawa, T., 1999. Factors influencing imitation of manipulatory actions in chimpanzees (*Pan troglodytes*). Journal of Computational Psychology 113, 128–136.

Narayanan, S.S., 1997. Knowledge-based action representations for metaphor and aspect (KARMA). thesis, University of California at Berkeley.

Oztop, E., Arbib, M.A., 2002. Schema design and implementation of the grasp-related mirror neuron system. Biological Cybernetics 87 (2), 116–140.

Oztop, E., Kawato, M., Arbib, M., 2006. Mirror neurons and imitation: a computationally guided review. Neural Networks 19 (3), 254–271.

Piaget, J., 1954. The Construction of Reality in a Child. Norton.

Piaget, J., 1971. Biology and Knowledge. Edinburgh University Press.

Pinker, S., Bloom, P., 1990. Natural Language and Natural Selection. Behavioral and Brain Sciences 13, 707–784.

Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. Trends in Neuroscience 21 (5), 188–194.

Rizzolatti, G., Berti, A., Gallese, V., 2000. Spatial neglect: neurophysiological bases, cortical circuits and theories, second ed.. In: Boller, F., Grafman, J., Rizzolatti, G. (Eds.), Handbook of Neuropsychology, vol. 1 Elsevier Science, pp. 503–537.

Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L., 1996. Premotor cortex and the recognition of motor actions. Cognitive Brain Research 3, 131–141.

Rizzolatti, G., Luppino, G., 2001. The cortical motor system. Neuron 31 (6), 889–901.

Rizzolatti, G., Luppino, G., 2003. Grasping movements: visuomotor transformations. In: Arbib, M.A. (Ed.), The Handbook of Brain Theory and Neural Networks, second ed. The MIT Press, pp. 501–504.

Rosch, E., 1978. Principles of categorization. In: Rosch, E., Lloyd, B.B. (Eds.), Cognition and Categorization. Laurence Erlbaum Associates, pp. 27–48.

Roy, A.C., Arbib, M.A., 2005. The syntactic motor system. Gesture 5, 7–37.

Seyfarth, R.M., Cheney, D.L., Bergman, T.J., 2005. Primate social cognition and the origins of language. Trends in Cognitive Sciences 9 (6), 264–266.
Shastri, L., Ajjanagadde, V., 1993. From simple associations to systematic reasoning. Behavioral and Brain Sciences 16, 417–494.
Sternberg, D.A., Christiansen, M.H., 2006. The implications of bilingualism and multilingualism on potential evolved language mechanisms. In: Proceedings of the 6th International Conference on the Evolution of Language, pp. 333–340.
Stokoe, W.C., 2001. Language in Hand: Why Sign Came Before Speech. Gallaudet University Press.
Talmy, L., 1985. Lexicalization patterns: semantic structure in lexical forms. In: Shoepen, T. (Ed.), Language Typology and Lexical Description: Grammatical Categories and the Lexicon. Cambridge University Press, pp. 36–149.
Tettamanti, M., Buccino, G., Saccuman, M.C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S.F., Perani, D., 2005. Listening to action-related sentences activates fronto-parietal motor circuits. Journal of Cognition Neuroscience 17 (2), 273–281.
Tomasello, M., 1999. The human adaptation for culture. Annual Reviews in Anthropology 28, 509–529.
Tomasello, M., 2003. Constructing a Language: A Usage-Based Theory of Language Acquisition. Harvard University Press.
Tomasello, M., Call, J., Warren, J., Frost, T., Carpenter, M., Nagell, K., 1997. The ontogeny of chimpanzee gestural signals. In: Wilcox, S., King, B., Steels, L. (Eds.), Evolution of Communication. John Benjamins Publishing Company, pp. 224–259.
Tomonaga, M., Tanaka, M., Matsuzawa, T., Myowa-Yamakoshi, M., Kosugi, D., Mizuno, Y., Okamoto, S., Yamaguchi, M.K., Bard, K., 2004. Development of social cognition in infant chimpanzees (*Pan troglodytes*): face recognition, smiling, gaze, and the lack of triadic interactions. Japanese Psychological Research 46, 227–235.
Umiltà, M.A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., Rizzolatti, G., 2001. I know what you are doing: a neurophysiological study. Neuron 31, 155–165.
Visalberghi, E., Fragaszy, D., 1990. Do monkeys ape? In: Parker, S., Gibson, K. (Eds.), Language and Intelligence in Monkeys and Apes: Comparative Developmental Perspectives. Cambridge University Press, pp. 247–273.
Voelkl, B., Huber, L., 2007. Imitation as faithful copying of a novel technique in marmoset monkeys. PLoS ONE 2, e611.
Webster, M.J., Bachevalier, J., Ungerleider, L.G., 1994. Connections of inferior temporal areas TEO and TE with parietal and frontal-cortex in macaque monkeys. Cerebral Cortex 4, 470–483.
Wohlschläger, A., Gattis, M., Bekkering, H., 2003. Action generation and action perception in imitation: an instance of the ideomotor principle. Philosophical Transactions in Royal Society London 358, 501–515.
Wolpert, D.M., Kawato, M., 1998. Multiple paired forward and inverse models for motor control. Neural Networks 11 (7–8), 1317–1329.
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: Social Function and the Origins of Linguistic Form. Cambridge University Press, pp. 202–285.