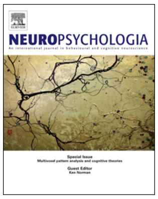
Neuropsychologia ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neuropsychologia

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



Language is handy but is it embodied?

Michael A. Arbib [a,b,c,*], Brad Gasser [b,c], Victor Barrès [b,c]

[a] Computer Science, University of Southern California, Los Angeles, CA 90089-2520, USA
[b] USC Brain Project, University of Southern California, Los Angeles, CA 90089-2520, USA
[c] Neuroscience Graduate Program, University of Southern California, Los Angeles, CA 90089-2520, USA

ARTICLE INFO

Keywords:
Marc Jeannerod
Language
Action-oriented perception
Embodiment
Mirror System Hypothesis
Evolution of language
Language-ready brain
Neuro-primatology

ABSTRACT

Part 1 provides Arbib's reflections on the influence of Marc Jeannerod on his career. Part 2 recalls the Mirror System Hypothesis (MSH) for the evolution of the language-ready brain, a theory which emphasizes the role of manual action in grounding language evolution, thus giving one meaning for “language is handy”. Part 3 then joins in current debates over the notion of whether or not language is embodied. Our overall argument is that embodiment is a graded rather than binary concept, and that embodiment provides the evolutionary and developmental core of concepts and language, but that the modern human brain supports abstraction processes that make embodiment little relevant in a wide range of language use. We urge that, rather than debate the extent of embodiment, attention should turn to the integration of empirical studies with computational modeling to delineate in detail processes of abstraction, generalization, metaphor and more, bridging between modeling of neural mechanisms in macaque that may be posited for the brain of the last monkey–human common ancestor (LCA-m) and computational modeling of human language processing. Part 4 suggests that variants of construction grammar are well-suited to the latter task.

1. A personal history (Arbib in relation to Jeannerod) [1]

1.1. Prehistory: the 1970s. Action-oriented perception, schemas & computational neurolinguistics

A major theme of Marc Jeannerod's research has been to place cognition and perception squarely in the context of action (Jeannerod, 1997 provides an integrated perspective), with special attention to the visual control of hand movements. My own path to linking action and perception began with “What the frog's eye tells the frog brain” (Lettvin, Maturana, McCulloch, & Pitts, 1959) which showed that the frog's retina extracted features relevant to the detection of prey and predators. Through this, I came to meet David Ingle, a neuroethologist who reported that, when confronted with two fly-like stimuli, the frog would in a few cases snap at “the average fly” rather than at one of the stimuli (Ingle, 1968). This led Rich Didday and myself to consider “What the Frog's Eye Tells the *Frog*,” how the brain could transform retinal patterns into adaptive courses of behavior, a program my group pursued under the banner of *Rana computatrix*, the frog that computes (see, for example, Arbib, 1987; Didday, 1970; Ewert & Arbib, 1989).

Crucially, Ingle emphasized that what we learned of action-oriented perception in frogs was relevant to understanding mammalian brains as well. The symposium *Locating and identifying: two modes of visual processing* combined the insights of Ingle, Schneider, Trevarthen and Held (1967). For example, Schneider's study of hamsters distinguished a “where” system in the superior colliculus from a “what” system in cortex that allowed the hamster's behavior to depend on visual patterns whose discrimination was beyond the frog's capabilities. An intriguing follow-up was Humphrey's (1970) demonstration that a monkey without visual cortex could nonetheless navigate on simple visual cues like well-lit contours though having lost visual perception (compare “blindsight” in humans).

These influences helped make *action-oriented perception* a key concept in *The Metaphorical Brain: An Introduction to Cybernetics as Artificial Intelligence and Brain Theory* (Arbib, 1972). Of particular relevance here is the following: “The animal perceives its environment to the extent that it is *prepared* to *interact* with it. … Perception of an object generally involves the gaining of access to [schemas] for controlling interaction with the object, rather than simply generating a “name” for the object … [L]anguage can best be understood as a device which refines an already complex system – [and] is to be explained as a ‘recently’ evolved refinement of an underlying ability to interact with the environment.” One might say

---

* Corresponding author.
E-mail addresses: arbib@usc.edu (M.A. Arbib), bgasser@usc.edu (B. Gasser), barres@usc.edu (V. Barrès).

[1] Portions of this paper are based on a talk given by Michael Arbib for the Symposium in Honor of Marc Jeannerod held on October 29–30, 2012 in Lyon, France. Section 1 thus prefaces the main arguments of the paper with an appreciation of Jeannerod's contributions to cognitive neuroscience. Marc Jeannerod died on July 1, 2011.

0028-3932/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004

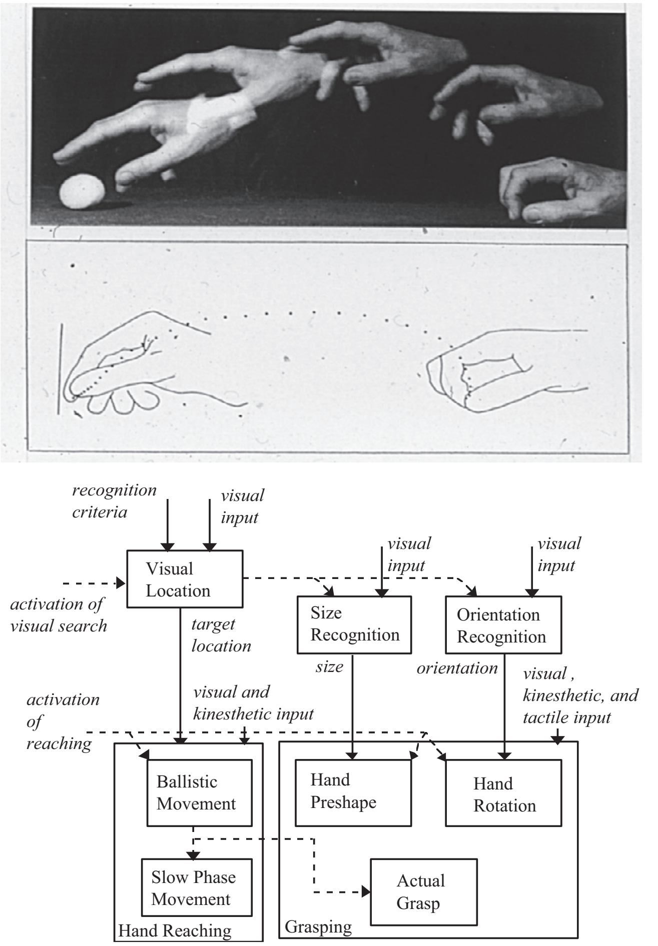


Fig. 1. (Top) (Upper) Preshaping of the hand while reaching to grasp; (Lower) Position of the thumb-tip traced from successive frames shows a fast initial movement followed by a slow completion of the grasp. (Courtesy of Marc Jeannerod. Adapted from Jeannerod & Biguer, 1982). (Bottom) A coordinated control program linking perceptual and motor schemas to represent this behavior. Solid lines show transfer of data; dashed lines show transfer of control. The transition from ballistic to slow reaching provides the control signal to initiate the enclose phase of the grasp. (Adapted from Arbib, 1981).

that my hypothesis was that language is rooted in embodiment and may modulate or be secondary to ongoing embodied behavior – but the argument still held that language also supported inferences and concepts that were abstract rather than embodied. One might know that President Nixon was a male by summoning a visual image with his five o'clock shadow, but most of us cannot summon an image of President Polk, and instead know he is male by "disembodied" inference from the generalization "All presidents of the United States have been male".

Another conceptual development came from seeking to reconcile working top-down from behavior with working bottom-up from neural circuitry, and forward from sensory receptors and back from muscles, describing the frog's visuomotor behavior in terms of the interaction of *perceptual schemas* and *motor schemas,* with *cooperative computation* (competition and cooperation based on activity levels) between schemas, rather than binary choices, underlying behavior. Cooperative computation of schemas was taken up by Allen Hanson and Ed Riseman in their VISIONS system for interpreting a visual scene – the result being a spatially anchored schema assemblage. A first-pass segmentation of the image provided the basis for invoking perceptual schemas for entities which represented visual correlates of entities like *sky, roof, house, wall,* and *grass* and possible spatial relations between them in New England suburban scenes. Competition and cooperation proceeded both bottom-up (aggregating visual features to instantiate a schema) and top-down (as instantiation of schemas to interpret one region provided cues to support or oppose interpretations for nearby regions) to yield an interpretation associating schemas with distinctive regions of the scene (Hanson & Riseman, 1978). Although implemented on a serial computer, the system revealed an essentially brain-like style of distributed computation. The HEARSAY system provided a similar, and near contemporaneous, computer system for speech understanding (Lesser, Fennel, Erman, & Reddy, 1975).

Following up on these various studies, I collaborated with the aphasiologist David Caplan to argue that "Neurolinguistics Must Be Computational" (Arbib & Caplan, 1979). We showed how schema models might provide the necessary intermediary between neurolinguistic analysis and utilization of the fruits of modern neuroanatomy and neurophysiology.

1.2. *Marc Jeannerod and the centrality of action*

It was thanks to the frog – and more specifically to David Ingle – that I first met Marc Jeannerod. This was at the NATO Advanced Study Institute on *Advances in the Analysis of Visual Behavior* that David co-organized with Richard Mansfield and Mel Goodale at Brandeis University in June of 1978. Jeannerod's talk "Visuomotor mechanisms in reaching within extra-personal space" (later published as Jeannerod and Biguer (1982)) opened up a whole new dimension of schema theory for me. His insights into the preshaping of the human hand (Fig. 1, top) led me to the notion of a *coordinated control program* (Fig. 1 bottom, adapted from Arbib (1981)). Perceptual schemas here serve not only to recognize objects (as in VISIONS) and their properties but also to pass parameters to motor schemas – as in visuomotor coordination in the frog.

At the same 1978 meeting, Ungerleider and Mishkin introduced their classic distinction between the *what* (ventral) and *where* (dorsal) streams in the monkey. In due course, Jeannerod, Decety, and Michel (1994), Jeannerod, Michel, and Prablanc (1984) and Goodale, Milner, Jakobson, and Carey (1991) developed a *related* analysis of the human reach-to-grasp where the ventral stream determines *what* an object is, and the dorsal stream determines *how* to grasp it. Three observations: (i) Schneider had discovered a *what* versus *where* distinction between cortex and midbrain in the hamster. (ii) Ungerleider and Mishkin related *what* versus *where* to inferotemporal versus parietal cortex for monkeys during a memory task based on spatial pattern versus location, respectively. (iii) By contrast, Jeannerod and Goodale et al. looked at the online use of visual information during reaching to grasp an object and then extended the involvement of the dorsal stream to a variety of parameters (not just *where* the object was located) related to *how* the action was performed, consistent with the data and model of Fig. 1.

The publication of the model of Fig. 1 gave Ian Darian-Smith the erroneous impression that I had some expertise in the neural control of hand movements, and he invited me to speak at the IUPS Satellite Symposium on *Hand Function and the Neocortex* in Melbourne, Australia, in August, 1983. This provided a great stimulus to develop such expertise (Arbib, Iberall, & Lyons, 1985; Iberall, Bingham, & Arbib, 1986). This in turn led to increasing interaction with Marc Jeannerod which included sending two of my students, Peter Dominey and Bruce Hoff, to work with Marc in Lyon. In particular, Bruce addressed new studies in Lyon (Paulignan, Jeannerod, MacKenzie, & Marteniuk, 1991; Paulignan, MacKenzie, Marteniuk, & Jeannerod, 1991) of human kinematics which studied perturbations of the reach to grasp in which either the size or location of the object was perturbed after the grasp was initiated. This contradicted the hypothesis in my original model that the first phase of the reach was ballistic, but led to models of the motor schemas as dynamic control systems combining feedback and feedforward, and with coupling between them (Hoff & Arbib, 1991, 1993).

Please cite this article as: Arbib, M. A., et al. Language is handy but is it embodied? *Neuropsychologia* (2013), http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004

Fig. 2. Marc and Jacqueline Jeannerod, Michael and Prue Arbib, Giacomo Rizzolatti and Hideo Sakata at the HFSP workshop on cognitive control of movements and actions, Hakone, Japan, on November 20, 1991. Note added in proof: Hideo Sakata died on October 4, 2013.

At a chance meeting with the neurophysiologist Hideo Sakata at IBM Japan's Symposium on Neuro-Computers in Oiso, Japan, November, 1988. Hideo told me that he, Marc, and Giacomo Rizzolatti were planning to submit a proposal for research collaboration to the newly formed Human Frontier Science Program. This conversation led to my joining the group. This collaboration engendered not only exciting research but also a deep friendship between the scientists and their wives (Fig. 2).

Where the collaboration with Marc focused on analysis of human behavior, Hideo and Giacomo focused on neural correlates in the brain of macaque monkeys. Sakata's lab demonstrated that neurons in AIP (anterior intraparietal sulcus) responded to vision of an object with activity that correlated with "how" the object was to be grasped (which we viewed as an instance of *affordances* in the sense of Gibson (1966)) whereas data from Rizzolatti's lab showed how neurons in the area of premotor cortex they labeled F5 coded something akin to motor schemas. The insights from the first stage of our collaboration were set forth in "Grasping objects: the cortical mechanisms of visuomotor transformation" (Jeannerod, Arbib, Rizzolatti, & Sakata, 1995). Around the same time, Marc published a highly influential paper (Jeannerod, 1994), arguing for a functional equivalence between motor imagery and motor preparation. He showed how object attributes are processed in different neural pathways depending on the kind of task the subject is performing, with a pragmatic representation activated in which object affordances are transformed into specific motor schemas.[2] The dorsal path is more concerned with the "how" of converting locally discernible affordances of an object into motor parameters for interacting with it, whereas the ventral path goes beyond "what" the object is to take context, task and more into account to determine which actions are to be executed.

But the most famous discovery made during the HFSP collaboration was that of mirror neurons, and it is interesting to see how Jeannerod (1994) talked of them shortly after their discovery but before they were labeled *mirror neurons* (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996):

> Rizzolatti and his group have described a class of neurons in the rostral part of the inferior premotor cortex, which fire prior to and during specific actions performed by the animal (e.g., picking up a food morsel with a precision grip). Neuron discharge is usually not conditional to the hand used, nor to the orientation of the grip, it relates to the fact that the monkey performs that particular action (Rizzolatti et al., 1988). Recently, these authors noticed that the same neurons also fire while the monkey observes the experimenter performing the same action (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992). According to the authors, "there is always a clear link between the effective observed movement and that executed by the monkey and, often, only movements of the experimenter identical to those controlled by a given neuron are able to activate it". This very striking result supports the idea of representing neurons as a common substrate for motor preparation and imagery.

There is much more that could be said in appreciation of the influence of Marc Jeannerod on placing cognition and perception clearly in the context of action and motor representations, but enough has been said to set the stage for what follows.

1.3. *Modeling the reach to grasp and its recognition*

Building on Fig. 1 and the experiments from the labs of Rizzolatti and Sakata, we developed the FARS (Fagg–Arbib–Rizzolatti–Sakata) model (Fagg & Arbib, 1998) of parieto-frontal interactions in the reach to grasp. It combined a dorsal *how* pathway transforming object affordances into motor schemas for the appropriate grasp with a ventral *what* path exploiting object recognition in temporal cortex to support prefrontal planning of behavior that could include specification of which affordance to exploit for the current task. Thus, this is not a "choose one path" model. We then modeled associative learning processes that could establish mirror-neuron responses (Oztop & Arbib, 2002), emphasizing recognition of the trajectory of a hand's motion toward an object, with the corollary that in many cases an action could be recognized well before its completion. Recognition is not the same as understanding – we hold that understanding requires the involvement of the ventral system – and this model does not take the ventral path into account. We recently demonstrated the conceptually crucial point that mirror neurons could play a role in self-monitoring of one's own actions to yield a more adaptive behavioral controller (augmented competitive queuing [ACQ] model, Bonaiuto & Arbib, 2010), a function which may have preceded the social role in evolution.

Buccino et al. (2004) used fMRI to study subjects viewing a video, without sound, in which individuals (man, monkey, or dog) bite or perform communicative acts. There was a clear overlap of the cortical areas that became active in watching man, monkey or dog biting, including activation in areas considered to be mirror systems. However, although the sight of a man moving his lips as if he were talking induced strong "mirror system" activation, the activation was weak when subjects watched monkey lip smacking, and disappeared when they watched the dog barking. Buccino et al. conclude that actions in the motor repertoire of the observer are mapped on the observer's motor system via mirror neurons, whereas actions that do not belong to this repertoire (e.g., barking) are recognized without such mapping. However, we suggest that the understanding of *all* actions involves ventral mechanisms which need not (but, obviously, can for actions in the observer's repertoire) involve the mirror system strongly.

2. Language is Handy

The phrase "Language is Handy" is intentionally ambiguous. On the one hand [sic], it means "Language is Useful" but here we follow the other hand to the claim that "Mechanisms for the visual

---

[2] Jeannerod used "pragmatic" in the sense for which we use the word "praxic," to distinguish *praxic* (practical) actions upon objects in the world from *communicative* actions. This differs from the linguist's use of the term *pragmatics* in counterpoint to *semantics*.

Please cite this article as: Arbib, M. A., et al. Language is handy but is it embodied? Neuropsychologia (2013), http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004

control of hand movements played a crucial role in the evolution of the human language-ready brain". The view that emerged from the article "Language Within Our Grasp" (Rizzolatti & Arbib, 1998) is an oft-told tale, so let's be brief; an extended presentation can be found in *How the Brain Got Language* (Arbib, 2012). (The book is the subject of a dozen commentaries in *Language and Cognition,* Volume 5, 2–3, 2013.) The key observations were that macaque premotor area F5 (a site of mirror neurons for manual actions) is homologous to human Broca's area, traditionally thought of in relation to speech production, and that there is "mirror system for grasping" in or near this area. Lesions akin to those yielding Broca's aphasia yield corresponding deficits in people who communicated using sign language (Poizner, Klima, & Bellugi, 1987) – thus letting us see Broca's area in relation both to manual praxic action[3] and to language considered in the manual as well as the vocal domain. This led us to hypothesize that the basis for *language parity* (the approximate matching of what the speaker intends and the hearer understands, Liberman & Mattingly, 1989) evolved *atop* the mirror system for grasping, rooting speech in communication based on manual gesture. Note the crucial word "atop" – monkeys have mirror neurons, but do not have language. Evolution had to build not only on the mirror system but also on systems "beyond the mirror" to yield the human language-ready brain.

2.1. *The mirror system hypothesis: an overview*

The developed version of the Mirror System Hypothesis (MSH) includes at least the following:

Starting with a mirror system for grasping shared with the last common ancestor of humans and monkeys (LCA-m), biological evolution of the *pre-hominid* brain (coupled with cultural evolution of practices in skill sharing and gestural communication) yielded a simple imitation system for manual actions shared with the common ancestor of humans and great apes (LCA-a). Further coupling of biological evolution with niche construction (Laland, Odling-Smee, & Feldman, 2000) during hominid evolution yielded an ancestral species whose brain supported a *complex imitation* system combining *complex action recognition* with the ability to use this recognition to attempt to approximate the performance on this basis (Arbib, 2002; Byrne & Russon, 1998).

This in turn set the stage for the emergence of a brain that made *intentional pantomime* possible. For the evolutionary hypothesis, it is crucial to distinguish this ancestral pantomime from the highly convention-laden productions of a game of charades. Rather, we refer to the *artless* sketching of an action to indicate either the action itself or something associated with it. A charade is parasitic on language, whereas "artless pantomime" is parasitic on the motor system: rather than acting on an object, one acts as if the object were present to convey something about the action, the object or some larger situation in which they are embedded, and which may thus elicit an appropriate response – dependent on context – from the observer for whom the pantomime is performed. The crucial point for MSH is that once the biological potential for pantomime was realized as a social practice, our ancestors could readily exapt complex action recognition to perform novel communicative actions to support an *open-ended semantics*. A continuing debate is thus with scholars who argue that language emerged as speech as a direct descendant of monkey-like vocalizations (e.g., MacNeilage & Davis, 2005) rather than indirectly via manual gesture (Armstrong, Stokoe, & Wilcox, 1995; Corballis, 2002; Hewes, 1973; Kimura, 1993). However, pantomime is costly to produce and highly ambiguous. MSH thus posits a transition from spontaneous adaptation of praxic skills for communication to a capability for *protosign,* the development by a community of ritualized gestures with relatively delimited referents, to support more reliable communication and thus opened the path towards language, with its capacity to express ideas far beyond the reach of pantomime.

An extended range of meanings (far more diverse than the vocalizations and gestures of other primates) was thus opened up by the expressive power of pantomime and then associated with increasingly conventionalized and non-pantomimic gestures as protosign developed. The further argument is that a relatively limited form of protosign (as distinct from a full sign language, linking grammar and lexicon) sufficed to create an adaptive pressure for the evolution of brain mechanisms for the increased voluntary control of vocal articulation, something lacking in other primates: elementary protosign provided the necessary scaffolding for the emergence of *protospeech*.[4]

Early protosign was a very distant ancestor of a modern sign language like American Sign Language (ASL), a "fully human" language endowed with a rich lexicon and grammar and an extended expressive capacity far transcending that of early protolanguages. We have argued (Arbib, 2008) that it is the availability of complex imitation that allows

- the original protowords to be *fractionated* to yield new words (an utterance that encodes *flying-bird* as a *holophrase* might over time come to be replaced by the concatenation of variant substrings coding *flying* and *bird* separately)
- and also to yield constructions which serve both to combine these emergent words in a (non-holophrastic) utterance as well as to combine "similar" words to express novel meanings.

Over time, both the lexicons and grammars (sets of constructions) of protolanguages would be enriched, to yield a spectrum of increasingly complex protolanguages which would eventually reach such complexity that we might better call them "simple languages" than "complex protolanguages". Implicit in this statement is the view that grammar is *not* all-or-none. Languages emerged from protolanguages through a process of *bricolage* (tinkering). This yielded, cumulatively, many novel words and constructions to handle special problems of communication. Many, but not all, constructions became more regularized, with general "rules" emerging both consciously and unconsciously only as generalizations could be imposed on a range of ad hoc mechanisms. The result: a spiraling co-evolution of communication and representation, extending the repertoire of describable actions, objects and situations – and supporting an expanding range of increasingly abstract concepts in the process.

2.2. *Back to the brain, briefly*

Fig. 3 builds on the distinction articulated in Section 1.2. The dorsal path is more concerned with the *how* of converting affordances of an object into motor parameters for interacting with it, whereas the ventral path goes beyond *what* the object is to take into account context, task and more to determine which actions are to be executed.

(1) The dorsal "Mirror for (Compound) Actions" is a system that can both generate an action appropriate to selected affordances

---

[3] See footnote 2.

[4] A contrary view, the *musical protolanguage hypothesis*, has been advanced by Mithen (2005) and Fitch (2010), building on Darwin's hypothesis that proto-song preceded language (Darwin, 1871).

Please cite this article as: Arbib, M. A., et al. Language is handy but is it embodied? *Neuropsychologia* (2013), http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004

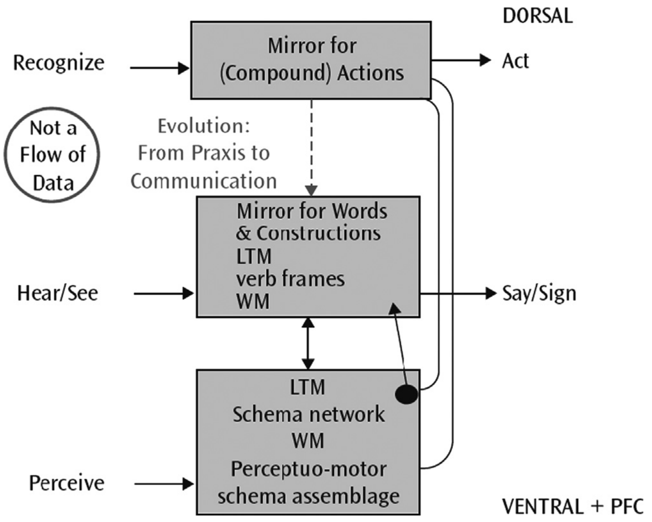


Fig. 3. A mirror and production system for praxic actions (and their aggregation into coordinated control systems) provides the evolutionary basis for the emergence of a mirror and production system for words and larger utterances *as articulatory structures*. The argument is that recognition of a word (from the "how" of its articulation) must be differentiated from interpreting the meaning of the word (its "what") via a ventral pathway.

and also recognize the form of an action being performed by another.

(2) The ventral "Schema Network" is a system that provides "understanding" beyond that accommodated by the dorsal stream, using perceptual schemas encoded in long term memories (LTM) to update a working memory (WM) of the current scene in relation to current goals.

(3) The dorsal "Mirror for Words and Constructions" evolved from the dorsal "Mirror for (Compound) Actions," but the *meaning* of words and utterances still resides in the ventral "Schema Network," which is now enriched by concepts made "thinkable" only by the new knowledge that language and culture make possible.

When we hear a familiar word, the ventral path alone may often suffice to recognize what the word "is" in the sense of being able to access its meaning and repeat the word in one's own accent, but the dorsal path is needed if one is to attempt to replicate the pronunciation of a word if it was produced in an unfamiliar accent other than one's own (Moulin-Frier & Arbib, 2013), or if one attempts to repeat a nonsense word whose articulatory form is not in one's repertoire.

2.3. Conversation considered

One of the methodological problems that plague the embodiment literature is disagreement over what "comprehension" means. Unfortunately, many studies accept the view that meanings are attached to words while syntax only stipulates how to properly compose them (strong compositionality), and that comprehension is an automatic translation of some semantic content of a sentence into some semantic representation in the brain. We instead endorse the view that human communication is not a mere coding–decoding procedure; people use context to guide interpretation (Arbib & Hesse, 1986; Dessalles, 2008; Sperber & Wilson, 1986/95). Sperber and Wilson (2002) advance the broadly Gricean view that pragmatic interpretation is ultimately an exercise in mind-reading, the ability to infer the mental states of others – so long as it is understood that the "inference" involved (cooperative computation) need not be conscious. However, they open a new can of worms when they argue that the interpretation process involves a dedicated "metacommunicative" comprehension module, with its own special principles and mechanisms. By contrast, we see the interpretation process as being strongly linked to other mechanisms rather than being encapsulated. In specific conditions motor activity takes part in the construction of sentence meaning, in others it does not. The issue remains of the "conscious readout" from a network of varied activation with the consequent "subthreshold thoughts" that may be "promoted" by associations with other material, whether perceived, recalled or fabricated.

The emphasis of MSH on bricolage and a related assessment of the role of cultural evolution in the diversity of human languages (Haspelmath, Dryer, Gil, & Comrie, 2005) run contrary to Chomsky's protean notion of an *innate* Universal Grammar (see Arbib, 2012, pp.47–52 for a critique) or the view that recursion is the key to language (see Arbib, 2012, pp. 168–9; contra Hauser, Chomsky, & Fitch, 2002). Indeed, we see many human actions and perceptual acts as exhibiting a recursive structure, and MSH shows how these may be "lifted" to language. *Parity* has a central place in MSH – but this role establishes the parity of sound and/or motor patterns via the dorsal pathway, with the linkage to meaning requiring a ventral contribution. By positing that "lifting" complex imitation from manual skill to a new, open-ended communicative domain, MSH offers an account of the piecemeal emergence of human intentional communication in Grice's sense that the uptake of communication requires an addressee to figure out the content of another's communicative intention (Grice, 1957, 1969, 1975). Praxic actions are performed to achieve a goal and to do so with little unnecessary effort. Thus, when we turn to communicative goals, MSH explains why the speaker/signer wishes to expend relatively little effort while reaching that goal of being understood by the listener/observer.

However, MSH has little to say about Grice's view that intentional communication is an exercise in mindreading. Complementing MSH, Jeannerod (2005) stated: "People generate intentions, they have goals, they feel emotions and affects. It is essential for each of us to penetrate the internal world of others, particularly when their intentions or goals are directed to us, or when their emotions relate to us. This very fact of knowing that one is the subject of others' mental states … is a critical condition for fully human communication between individuals." He contrasts the *narrative self*, where verbalization or imagination is based on memories retrieved via declarative memory systems which provide a strong feeling of continuity in our conscious experience, with the *embodied self* which operates on a moment to moment basis whereby one becomes conscious, if only transiently, of one's current bodily state.

Where Jeannerod here links *embodied* to "what the body is doing now," other scholars would argue that even the narrative self, indeed language in general, is *embodied* in the sense that our use of words acquires its meaning in part by awakening echoes of embodied experience in the brains and bodies of the speaker and hearer. Gallese and Lakoff (2005) argued that "conceptual knowledge is *embodied*, … mapped within our sensory-motor system". For them: (a) Imagining is the *mental simulation* of action or perception, using many of the same neurons as in actual acting or perceiving. (See Jacob and Jeannerod (2005) for a critique of this rather broad use of the term simulation.) (b) Understanding *is* imagination. 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. In the same vein, Gallese, Keysers, and Rizzolatti (2004) posit that "mirror mechanisms … allow us to directly understand the meaning of the actions and emotions of others by internally

Please cite this article as: Arbib, M. A., et al. Language is handy but is it embodied? *Neuropsychologia* (2013), http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004

replicating ('simulating') them without any … conceptual reasoning". In the spirit of Fig. 3 and the quote from Jeannerod (2005), we find it more plausible that, for mind reading and social cognition, mirror mechanisms must be complemented by processes that depend in part on language but also require an attitude to social interaction resulting from as subtle a mix of biological and cultural evolution as that offered by MSH. Studies by Tomasello and colleagues (e.g., Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012) may usefully complement MSH, especially if neural correlates of social dominance in monkeys (e.g., Fujii, Hihara, & Iriki, 2008; Fujii, Hihara, Nagasaka, & Iriki, 2008; Santos, Nagasaka, Fujii, & Nakahara, 2011) are factored into the evolutionary hypotheses.

Jeannerod (2005) concludes his account with a figure sketching the interaction between two agents. "Each agent builds in his brain a representation of both his own intended actions … and the potential actions of the other agent with whom he interacts. These partly overlapping representations are used by each agent to build a set of predictions and estimates about the social consequences of the represented actions, if and when they would be executed. When an action comes to execution, it is perceived by the other agent as a set of social signals which do or do not confirm his predictions and possibly modify his beliefs and desires. (Jeannerod, 2005, caption of his Fig. 1)" At a finer grain of analysis, Pickering & Garrod (2013a, 2013b), presenting production and comprehension of language as interwoven, import the concept of forward models from motor control (Wolpert & Kawato, 1998) to stress that these processes are united by the ability of people to predict themselves and each other. (See also Wolpert, Doya, & Kawato, 2003.)

The crucial point is that we must increasingly focus on the brains of interacting agents (we start with *dyadic* brain simulation in Section 3.2). The brain involves many different pathways which compete and cooperate, and so the relevant principles of cooperative computation may also help illuminate the neural mechanisms of conversation, or of social interaction more generally.

3. Is language embodied?

There are now many different takes on what embodiment means in relation to concepts and language (see, for instance, the papers by Jeannerod and by Gallese and Goldman cited earlier, as well as Clark (2008), de Bruin and Gallagher (2012), Gallese and Sinigaglia (2011), Goldman (2012), Goldman and de Vignemont (2009) and Jacob (2012), just for starters). Rather than reviewing this literature in detail, we will offer just enough background to make explicit how we approach the challenge of modeling brain mechanisms supporting language in a fashion consistent with our early mantra (Arbib, 1972) that "[L]anguage can best be understood as a device which refines an already complex system – [and] is to be explained as a 'recently' evolved refinement of an underlying ability to interact with the environment". Language cannot be subsumed in a generic primate *Bauplan* for sensorimotor interaction.

3.1. A view of embodiment

What does it mean to claim "Language is embodied"? If all it meant were that "The use of language is a human ability; humans have bodies; therefore language is embodied" then further discussion is pointless. We want instead to have an approach that leaves open to discussion whether *all* of a given language is embodied, *some* of a language is embodied, or *none* of a language is embodied. The "inner workings" of language are to be distinguished from the sensorimotor periphery of language use. Consider the meaning of *kick*. Its translation in American Sign Language (ASL) would have the same semantics (at least to a first approximation) as in English, and both forms refer to the same type of embodied action, but their *use* involves distinct embodiments (ear–voice versus eye–hand). Moreover, recognition/performance of the kicking action is only part of the semantics of "kick", so to say "the underlying concept is in part embodied" does not say "the 'entire' concept is embodied". And when we turn to a phrase like "differential equation," the embodiment of perception and production remains, but any embodiment of the concept is rarely essential to its meaning. From our cooperative computation perspective, very different sets of schemas may be deployed on different occasions (Arbib & Hesse, 1986). Cooperative computation is of the essence: concepts live in a schema network. There is no "full" semantics; rather, there are diverse context-dependent activations.

Wilson (2002) has characterized six different conceptual views of embodied cognition. The third, *cognition is for action* recalls the action-oriented perspective we take *as a starting point*. It is frustrating to see Engel, Maye, Kurthen, and König (2013) assert that "an action-oriented paradigm is emerging which was earliest and most explicitly developed in robotics [with their earliest citation from 1986] and more recently began to impact on cognitive psychology and neurobiology," citing Jeannerod (2001) as if he were a follower and ignoring the pioneering efforts from the 1970s and earlier sampled in Sections 1 and 2.

Strong embodiment views consider that simulation of sensorimotor schemas provides the sole and ultimate support of understanding (Feldman, 2010; Gallese & Lakoff, 2005; Pulvermüller, 2005). *Weak embodiment* or hybrid views simply claim that motor/modal simulations can enrich and/or change the phenomenological quality of comprehension but are not necessary (Binder & Desai, 2011; Dove, 2010; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2010). Conversely, *disembodied* perspectives claim that any involvement of the sensori-motor system is epiphenomenal to the language comprehension process in which it plays no direct causal role (Mahon & Caramazza, 2008). Clearly, the human brain's capacity for language implicates the importance of mechanisms beyond the general primate capacity for interaction with the environment. The debate should really reduce, then, to whether these novel mechanisms can best be understood in relation to other mechanisms for action and perception or as independent systems suited only for symbol processing. Only by looking at the language comprehension of modern humans from the point of view of brain theory and language evolution will it be possible to rigorously delineate the way diverse brain systems contribute in varying combinations to a whole range of language tasks.

Mahon and Caramazza's (2008) "domain-specific hypothesis" contends that neural systems provide important (innate) constraints on the organization of conceptual knowledge, and that although abstract conceptual knowledge includes a "conduit" that can interface with purely embodied action in the world, the larger expressive power that abstract concepts provide goes beyond the limits of the "body" to act in more flexible ways. As an example, they interpret data showing "automatic" activation of sensory and/or motor information during linguistic tasks as ancillary to understanding, while claiming that "stronger" versions of embodiment imply such activation to be necessary in retrieving semantic content – compare the Perceptual Symbol Systems of Barsalou (1999).

Pulvermüller (2013) places sensorimotor systems in a more obligatory role in semantic knowledge. He contends that sensorimotor systems do more than "color" concepts and knowledge, but necessarily are involved in the full semantic content of some concepts, and that without these systems any particular concept may be "diluted," as in lesioned or neurodegenerative patients (Kemmerer, Rudrauf, Manzel, & Tranel, 2012). Pulvermüller

Please cite this article as: Arbib, M. A., et al. Language is handy but is it embodied? *Neuropsychologia* (2013), http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004

additionally cites transcranial magnetic stimulation (TMS) studies that suggest experimental modulation of motor centers can either facilitate or interfere with perception of phonetic stimuli, depending on somatotopy (D'Ausilio et al., 2009), with similar results at the word level (Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005). He cites computational modeling work which show Hebbian learning can suffice for sensorimotor networks to support word learning (Garagnani, Wennekers, & Pulvermüller, 2008), but this study is not germane to the issue at hand. The model establishes bidirectional pathways between neural codes for the auditory form of a word and the articulatory form of the word, but completely ignores the issue of how these pathways are linked to embodied activity or other semantic forms. It thus addresses the dorsal, but not the ventral, path in Fig. 3.

The observation that the motor expression of a word (the signifier) may bear no resemblance to a praxic action is still to be distinguished from the question of whether or not its meaning (the signified) must be understood as essentially embodied. For example, one might argue that any digit from 1 through 9 is embodied because we have a rich experience of counting on our fingers, and then argue that adding a new digit d at the end of a string w to represent the number $10[w]+[d]$ (where the [ ] indicates "number represented by") is an embodied operation, and yet still feel uncomfortable with the idea that

$$1,798,334,667,027,145,209,870$$

is really "embodied". This suggests that *embodiment* is a graded, not a binary, concept. Consider the following:

Basis Step: concepts which describe the objects we interact with, the actions we perform, and the emotions we feel are all embodied, as are the words which signify them.[5]

Induction Step: a word or phrase composed of embodied content and function words is itself embodied.

Conclusion: all language is embodied.

Here the problems run deeper. Consider this sentence, *Here the problems run deeper.* "Here" is somewhat abstracted from the embodied notion of "close to the speaker". The noun "problem" is extremely abstract, presumably generalized across a whole range of experiences where the path [sic] to a goal is not apparent. "Run deeper" is a metaphorical extension of the running of a river, where the river provides a metaphor for the flow [sic] of the argument, and the depth signifies aspects of the argument that have not "surfaced" in the preceding exposition. Finally, the attribution of "running" to a river is a metaphor abstracted from the fast forward motion of our embodied experience of running [at last]. To make the point further. "Here the problems run deeper" could be paraphrased as "But this raises tough problems". We have switched from "deep" to "raises," ignored metaphorical rivers and used a word whose current meaning is far from the (perhaps) original embodied experience of chewing on a tough piece of meat. Language thus appears to involve mechanisms which support the abstract use of words in a way which, here, is only muddied (back to that river?) by an appeal to embodiment.

The "argument" by induction fails not simply because *embodiment* is not a binary concept, but also because the basis step is inadequate (many words do not describe the objects we interact with, the actions we perform, or the emotions we feel). New words and concepts are formed not only by composition, but by processes of metaphor and generalization each of which may bleach out some degree of embodiment the concepts may have had. The recognition that embodiment is a graded rather than binary concept (on which many would agree) frees us to work *towards* an evolutionary brain theory that, for example, goes beyond sensorimotor mechanisms to support the use of metaphor. To address these challenges we will "build the bridge" from action to language from both ends (Fig. 4). Section 1.3 presented models of the reach to grasp and its recognition "from the monkey (more precisely, LCA-m) end" and Section 3.2 will extend this arc to include the ontogenetic ritualization of novel communicative gestures by apes. Section 4 will sample efforts to develop a computational framework for language use and learning based on some variation of construction grammar, thus beginning to build the bridge from the other end. Both the evolution and individual acquisition of language have roots in embodied interaction with the world, but a brain that supports the full range of language must incorporate new mechanisms which can exploit semantic abstractions without necessary recourse to any links to embodiment. One is reminded of Wittgenstein's ladder which, once climbed, may be discarded (Wittgenstein, 1921), or of the scaffolding of a building, a metaphor involved in the claim that early protosign provided the scaffolding for the emergence of protospeech.

Fig. 4. The Sydney Harbour Bridge under construction from both ends (Photo credit: National Library of Australia, nal.pic-vn3071984-v).

In any case, the question remains, "What shall we mean by *embodied*?" We make a Jeannerodian move, namely to relate embodiment to an *action repertoire and the perceptions which serve it*. More than mere linkage to a perceptual or motor modality, the claim becomes that embodiment is *essentially* linked to the way the animal's body engages with both the external world and the internal world of visceromotor underpinnings of motivation and emotion. Thus the nature of embodiment for a frog or a tick is very different from the nature of embodiment for a human. A major problem, though, is that human technology is constantly changing the boundaries of the human *Umwelt*. Consider the Higgs boson. It required the highly complex mathematics of the Standard Model of particle physics to determine what to look for, plus almost 40 years of effort, including construction of one of the most complex experimental facilities ever built, the Large Hadron Collider. Is "the Higgs boson," an embodied concept? Is a sentence that discusses it embodied? Our answer is "no". It is too many "induction steps" away from what we can do with our bare hands and naked eye. Yet, surprisingly, we may accept "unicorn" as an embodied concept, even though mythological, since we may build on embodied experiences of riding a horse and of holding a horn of some kind – what one might call "embodied imagination". Indeed, one of the many virtues of language is that it may alert one to how to behave

---

[5] Cognitive linguists have demonstrated that the root meaning of many function words, and not just content words, can be viewed as embodied. For example, *in* can be linked to a movement from outside to inside a container, while *this* versus *that* can be related to proximity or otherwise to the speaker's body (Langacker, 1986; Talmy, 1988).

Please cite this article as: Arbib, M. A., et al. Language is handy but is it embodied? *Neuropsychologia* (2013), http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004

in situations that one has not experienced. “Don't pet the tiger; he might try to eat you.” However, there is no guarantee that even something expressed at the “embodiment end” of language need be true – or that the concept of “truth” is embodied.

3.2. Biological evolution through the lens of comparative (neuro) primatology

Apes, as compared to macaques, appear to be more skilled at learning from others' performances of particular actions both in laboratory designs (Dean, Kendal, Schapiro, Thierry, & Laland, 2012) and in the wild (Byrne & Russon, 1998). Variation across geographical sites in the behavior of chimpanzees, for example, has led some to speak of ape “cultures” which persist because of these learning processes (Whiten et al., 1999) even though they are qualitatively simpler than human cultures. Yet all these primates exhibit great dexterity as well as patterns of social interaction that show appreciation of at least the standing of individuals within a dominance hierarchy. They share a similar embodiment considered as a primate *Bauplan* that establishes a core action repertoire and the perceptions that serve it. Thus, at a minimum, one might argue for embodiment in two different terms: (a) the increase in human dexterity and bipedal locomotion as well as other changes in the *Homo sapiens* body plan opens up a range of embodied experience denied other primates (while perhaps foreclosing others); while (b) the development of neural mechanisms for communicative control of hands, face and vocal apparatus makes an immensely wider range of embodied experience expressible in explicit, intended communication.

Studies of ape communication anchor hypotheses about the neural, behavioral and communicative repertoire of LCA-a. It is currently unclear whether and to what extent apes acquire particular gestural forms through social interaction rather than learning to draw their expression from some innate “gestural space” (Call & Tomasello, 2007; Hobaiter & Byrne, 2011; Perlman, Tanner, & King, 2012). Still, both views would seem to demand a treatment of the socially constrained learning processes involved in the competent production and comprehension of the manual and vocal gestures. How are their expressions contextualized; how are they organized neurally; why are different patterns of use seen in the same community, and thus how would interaction in a physical and social world tune their communicative behaviors?

Elsewhere (Gasser, Cartmill, & Arbib, 2013), we addressed the notion of *ontogenetic ritualization* (Call & Tomasello, 2007) to provide a conceptual model of how dyads of apes may generate a gestural form because one influences the behavior of the other through mutual interaction which yields a truncated version of a larger instrumental action sequence originally intended to *physically* exert that influence. A computational version (Arbib, Ghanesh, & Gasser, in press) instantiates dyadic brain modeling, building on the ACQ model of Section 1.3. We model how interactions between agents with similar brains may differentially adapt those brains over time. This allows us to assess which basic sensori-motor processes are needed to learn communicative signals from interaction in physical and social worlds. Importantly, we seek to delineate brain mechanisms distinguishing what apes and monkeys can learn. In particular, we assessed comparative data on fiber tract differences in primates that suggest important evolutionary changes in kinematic processing of others' action as we move from LCA-m, through LCA-a (Hecht et al., 2013). In this way, our model of ape gestural learning goes beyond “primate-general” circuitry. This sets the stage for future modeling contrasting ape and human brains to crystallize debate about what supports the language-readiness of the human brain. One might conclude from comparative studies such as that of Rilling et al. (2008) that “a bigger arcuate fasciculus implies language” but this sidesteps the computational questions of what the distinctively human arcuate adds beyond bandwidth. Our challenge is to use comparative neuroprimatology to extract clues for deepening our ongoing attempt to understand the evolution of the human language-ready brain. MSH posits that pantomime, enriched by deictic gestures which foreshadow the demonstratives of language (Diessel, 2013) provides the avenue to escape this restriction in communicative behavior of non-humans by drawing attention to details of the pantomimed “scene” beyond the instrumental action, and allow attention then to be drawn to “declarative” propositions. But this is a very early step toward the abstractions that can be expressed within modern cultures.

3.3. Cultural evolution from protolanguages to languages

A human signer may have aphasia, and thus lose the ability to employ the agreed on signs and syntax of his Deaf community, and yet be able to pantomime – demonstrating that (proto)sign involves brain systems beyond those adequate for praxic action and for caricaturing such actions to extract communicative actions on an ad hoc basis (i.e., non-conventionalized; not sanctioned by community use). One subject who could not sign *flying* in ASL could extend his arms and move in a fashion reminiscent of how a plane might bank and turn (Corina et al., 1992). Is this pantomime embodied? Well, it employs movements of the body – but it does not correspond to any human effectors but rather maps the spatial structure and motion of a moving object onto the panto-mimic's body. This seems to be an interestingly transitional case where one might say that a concept is “quasi-embodied” to the extent that one can imagine oneself as an object or in a body to which that concept applies. Nonetheless, the performance does not include the core knowledge that flying involves movement through the air that continues for a long time without support from the ground.

As human cultures – and the “need” to communicate about more complex ideas, and with more nuance – became a greater selective pressure, complex constructions and gesturing or signing techniques – and corresponding cognitive changes – would be needed to successfully coordinate this transfer of knowledge, and to fuel the growth of languages which could increasingly express ideas and situations with little or no direct links to embodiment.

3.4. Development of the individual child

The child first acquiring her mother tongue is experiencing the words as embodied in the sense that they bid others to meet the child's bodily needs. However, as the child matures, the process of abstraction yields concepts and words and constructions and utterances that may not be embodied, even though they are encoded as schemas in an embodied brain. Piaget (1954) charts four (or more) stages in the development of the child: Sensorimotor → Preoperational → Concrete Operational → Formal Operational. We regard this as a high-level view of the cumulative effect of multiple small changes punctuated by various “phase transitions”. The community in which a child is immersed gives her access to skills, concepts, vocabulary and constructions that would otherwise be unattainable, resting as they do on a long process of historical accumulation. The human never discards the sensorimotor stage but rather, through a process of generalization, automatization and hierarchical combination continually expands and enriches the repertoire of sensory and motor schemas through their embedding in a far larger network of more abstract knowledge – much of which, we claim, is not usefully thought of as embodied. With each generation, a community of speakers using a language – a “brain-ready” language – may dilute the coupling between certain linguistic concepts and the embodied experiences

Please cite this article as: Arbib, M. A., et al. Language is handy but is it embodied? *Neuropsychologia* (2013), http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004

that may have been at the very root of their etymology. Consider the following extract from Pulvermüller (2013):

> How would we explain the meaning of "justice", say to a child? Typically by mentioning situations that can be taken as instantiations of JUSTICE – children receiving each the same amount of sweets, a thief having to pay for stolen goods, a killer being locked away for life for his killings. … Crucially, without concrete examples of situational instantiations, or *action–perception manifestations,* it is hard to see how an abstract meaning or concept could be introduced to a language learner lacking similar concepts in the first place.

We agree that a concept must be acquired, and that links to specific examples may remain available. But a whole tower of abstraction must be built which the above excerpt glosses over. Consider "a thief having to pay for stolen goods". This rests on notions of ownership of property, of theft, of social compulsion ("have to") and of payment (through imprisonment or a fine, involving concepts of freedom and money, respectively). These are very abstract notions based on an understanding of a range of legal concepts in a given society, rather than *action–perception manifestations*. The issue then is: What must a brain possess to be capable of acquiring such concepts? This remains an open and challenging question.

Embodied representations can be *strategically recruited* in order to foster better understanding. Campisi and Özyürek (2013) analyzed the role that iconic co-speech gestures play while an adult is teaching a child new knowledge. They demonstrate how "iconicity in gestures can be a powerful communicative strategy in teaching new knowledge to children in demonstrations [and this] is in line with claims that it can be used as a scaffolding device in grounding knowledge in experience". Similarly, Kita and Davies (2009) report that adults' use of representational co-speech gestures increase with the difficulty of what they wish to convey (in this case describing geometric figures). Marghetis and Núñez (2013) analyzed hand gestures performed by subjects engaged in mathematical reasoning (carrying out a mathematical demonstration) highlighting how the mathematical concepts they employed are rooted in dynamic perceptuo-motor representations. However, even when the subtleties of higher mathematics are grounded in embodied experience, such experience may be refined or even contradicted when one seeks to verify propositions through a formal proof – an infinite-dimensional Hilbert space is a very different creature from 3-dimensional Euclidean space. Thus, while embodied concepts can serve as scaffolding for building abstraction either during development or when one is acquiring knowledge in a new field, *the scaffolding is not to be confused with the building*.

3.5. *Action, perception and language in the adult*

Generically, we may assess the role of "embodied" systems in the acquisition of a concept, yet find that, eventually, much (but not necessarily all – it depends on the concept and the context) future (competent) employment is accomplished without engagement of these systems. The brain operates according to *cooperative computation* (competition and cooperation based on activity levels) between schemas, rather than binary choices (Section 1.1). When we analyze multiple paths in the brain relevant to a particular function (as in the FARS model of Section 1.3) we do not assert that the brain operates in a "choose one path" mode; rather the paths cooperate to determine the overall cognitive state (whether purely mental, or yielding a course of overt action). However, in one task a path may not play any role in the final outcome, whereas in another task that same path may provide the crucial data that tips the balance in the cooperative computation in other pathways. Thus in some circumstances and for some concepts, an embodied skeletomotor or visceromotor concomitant may play the deciding role in a cognitive outcome, whereas in other cases, it may be irrelevant. Moreover, our discussion of "graded embodiment" makes clear that there are some concepts for which embodiment may often play a role, and others for which embodiment is seldom if ever relevant once the concept has been acquired.

Proponents of so-called "embodied cognition" cite data from neuroimaging and/or trans-cranial magnetic stimulation (TMS) showing that "motor" or "perceptual" (as the case may be) brain regions show activations different from baseline levels during *certain* language or action comprehension tasks, and thus claim that these regions assist in comprehension or understanding of the observed behaviors of others (Aziz-Zadeh & Damasio, 2008; Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni, 2006; Feldman & Narayanan, 2004; Pulvermüller & Fadiga, 2010). We stress that word "certain." For example, "semantic somatotopy", the activation of motor and premotor cortex according to the body-related action meaning of words has been established for verbs like *kicking, picking or licking* that refer to actions normally involving the legs, hands or tongue, respectively (Hauk, Johnsrude, & Pulvermüller, 2004). But is this more than "priming"? We know there are paths from word areas to motor areas since otherwise we could not respond to commands. However, this connection does not exhaust "the full semantics" of the verb. Being a foot action does not exhaust the motor realization of *kick* – but this objection falls once one accepts that brain activation may reflect a more specific kick-related activation of premotor cortex than fMRI can reveal. A more telling observation is that it appears implausible that premotor cortex encodes the notion that kicking someone hurts them, or that it is unwise to kick a ball in the direction of a window. Here we part ways from linguists who see a clean separation between semantics and pragmatics. They would object that hurting someone is not part of the meaning of *kick*. Nonetheless, a theory of language in the broad sense, which is what we want to understand (and recall our brief mention of Grice), should address why (2b) seems more appropriate than (2a) in response to (1) in a conversation.

(1) When Harry kicked that football, it broke the window next door.
(2a) Good for him!
(2b) Oh oh, he's in trouble!

But, of course, we can imagine situations in which the response might be (2a) – e.g., the second speaker hates the neighbor or (not shown in the words alone) the sentence is pronounced with heavy sarcasm. In the long run, we wish to understand the processes involved here. Our claim, then, is that language processing is heavily affected by world knowledge, but that much of that relevant knowledge is "bleached of embodiment". Nonetheless, in a dialog like

(3a) Harry kicked the ball, and discovered it was a cannonball.
(3b) Ow, that must have hurt!

embodiment might play a crucial role in generating (3b), though it is not clear that the somatotopy of the particular action involving forceful contact is at all relevant to the response.

In Section 2, we studied MSH to see one sense in which "language is handy" – MSH provides an evolutionary path from the generation and recognition of hand movements to language. It is also important to point out that each language is *useful* – a *handy* tool to act more flexibly in a cultural milieu in which that language is shared. Social cognition has been a large driver in behavior and brain evolution in primates (Byrne & Whiten, 1988; Cheney & Seyfarth, 2007; Deaner, Khera, & Platt, 2005; Dunbar, 1998; Seyfarth, Cheney, & Bergman, 2005). But while non-human primates

Please cite this article as: Arbib, M. A., et al. Language is handy but is it embodied? *Neuropsychologia* (2013), http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004

communicate about instrumental, "here-and-now" intentions, humans have the capacity to recall the past, think about the future, and consider alternative, counter-factual events. A robust theory of language evolution, and future variations on MSH, must consider these additional dimensions of language and cognition in humans (Suddendorf & Corballis, 2007).

We devote the rest of this section to a few empirical data linking sentence comprehension with varying degrees of "embodiment". First, consider negation. Tomasino, Weiss, and Fink (2010) examined hand-related motor and premotor activity associated with the comprehension of both positive imperatives and negative imperatives involving hand action and found that negative imperatives result in less BOLD activation in motor regions than positive imperatives. This makes sense if one considers a simple imperative like "Kick the ball" versus "Do not kick the ball" – one understands the meaning of *kick* in both cases, but the negation in the second command serves to inhibit motor priming. However, no such motor inhibition is involved in the statement "Los Angeles is not the capital of California". To address the timing of negation effects, Aravena et al. (2012) asked subjects to maintain a precision grip while listening to positive or negated action sentences. By measuring the variation of grip-force online during comprehension, the authors were able to show that positive sentences resulted in a significant increase in grip-force as early as 300 ms after presentation of the target word, an effect that is absent when action is negated. This suggests that the impact of negation on motor responses associated with comprehension of action-related words is fast. Kaup, Yaxley, Madden, Zwaan, and Lüdtke (2007) had argued that the comprehension of a negated state of affair required first the simulation of the positive state of affair which is later negated. However, the above data suggest that the involvement of sensory-motor systems is *context-dependent*.

Quantifiers are classically associated with many semantic problems (in the linguistic sense). Lidz, Pietroski, Halberda, and Hunter (2011) briefly presented subjects with a random number of colored dots and asked them to evaluate sentences such as "most of the dots are blue". Subjects tended to compare blue and non-blue dots as a verification procedure. Indeed, one can selectively attend at once to all the blue dots among a set that contains dots of many different colors, but one cannot attend selectively at once to the yellow and the green dots while ignoring the blue dots (Treisman & Souther, 1985). In this task, meaning of the quantifier "most" can be linked with properties of the visual system. But if we assert "most people in the world were born since 1980", the actual procedures required to verify or refute the claim would be very different, and the "disembodied" idea of very large numbers would be predominant.

A growing number of psycholinguistic studies use the Visual World Paradigm, in which eye-movements of listeners are recorded while they are both listening to utterances and looking at a visual scene. Assessing a picture need not be an embodied act (other than the peripheral role of eye movements) – as in distinguishing a Rembrandt from a Vermeer. Consider, then, expanding the notion of visual scene from a computer screen image to a *scene with which subjects can physically interact.* Chambers, Tanenhaus, and Magnuson (2004) had subjects face a table on which four objects were displayed. Eye-movements were monitored as the subjects hear sentences that contain a request to perform a given action, e.g. "Put the whistle on the folder in the box". Before hearing "in the box", the sentence is semantically ambiguous since it can either refer to the whistle that is on the folder or to another whistle. Referential ambiguity and its impact on eye movements have been the topic of many visual world paradigm studies, but Chambers et al. introduced a novel factor: before hearing the instruction sentence, the subjects were given a tool with which they were supposed to perform the action (e.g. a small hook). Using eye-tracking, the authors were able to show that the compatibility of the tool (e.g. the small hook) and the affordance of one of the potential referents (e.g. the string attached to only one of the whistles) constrained the referential domain relevant to syntactic processing. However, by stressing that a grammatical construction can usefully be studied in sentences whose meaning is embodied, we are not conceding that they only operate within embodied contexts.

Grossman et al. (2008) studied patients with amyotrophic lateral sclerosis and found that atrophy of the motor cortex hinders comprehension of action verbs but not of nouns. Arévalo, Baldo, and Dronkers (2012) asked patients who had had left-hemisphere strokes whether a given action word matched the action depicted in a picture or not. No correlation was found between the type of action (what body part it involves) and lesion to body-part-related motor areas. Papeo, Negri, Zadini, and Rumiati (2010) reported a double-dissociation in patients with left-hemisphere strokes between the capacity to pantomime an action and the capacity to produce and understand the verb that refers to it. Overall, these results seem to point towards a somewhat optional role of the motor system in language comprehension – but, given our overall framework of multi-path competition and cooperation, this does not rule out that this role may be crucial in certain circumstances.

4. Construction grammar: the other end of the bridge

Fig. 4 illustrated the metaphor of spanning from action to language by building a bridge from both ends. Sections 1.3 and 3.2 introduced efforts to build "from the monkey (more precisely, LCA-m) end". We now begin to build the bridge "from the other end," starting with a review of several efforts to develop a computational framework for language use and learning based on some variation of construction grammar. These approaches share the general assumptions of cognitive linguistics that language needs to be understood in terms of its use by a society of agents with situated bodies, that language processing should not be studied in isolation from what is known of sensori-motor systems, and that syntactic, semantic, and pragmatic aspects of language cannot be properly analyzed as separate components. *Construction grammars* are characterized, then, by a set of constructions, each carrying its own idiosyncratic mapping between form and meaning (for an introduction, see Croft and Cruse (2004)).

The so-called *Neural Theory of Language* (Feldman & Narayanan, 2004; but it is not really linked to neuroscience) has at its core a variant of motor schemas called X-Schemas (Narayanan, 1999). These were developed to computationally package some aspects of an action into a limited set of parameters that can be used either to direct action in the world or to carry out off-line simulations that, on their account, form the basis of language comprehension. Narayanan's system combined

(1) abstract world knowledge about a target domain (in this case, knowledge of international economics coded as a Belief Network),
(2) sensory-motor knowledge represented as a network of X-schemas, and
(3) metaphorical mappings between the two, linking belief values to X-Schema parameters.

He showed that this system could generate correct inferences as to the meaning of certain metaphors. For example, when presented with a newspaper headline such as "Liberalization plan stumbling", it concluded that there is an ongoing economic plan,

Please cite this article as: Arbib, M. A., et al. Language is handy but is it embodied? *Neuropsychologia* (2013), http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004

that it is facing difficulties, and that it is likely to fail. Such inferences are possible because the system can use X-Schemas to simulate the effect of stumbling on a *walk*-schema and map the resulting state (*falling* unless a lot of force is applied) to the concept of *difficulty* and *failure* in the target domain of economic policy. Building on this, *Embodied Construction Grammar* (ECG) sought to explain language comprehension in terms of sensory-motor simulations whereby linguistic meaning can be directly anchored in concrete action sentences. For abstract sentences, the claim is that the pervasive use of metaphorical mappings from an abstract target domain onto an embodied source domain still makes such simulations possible (Bergen & Chang, 2005; Feldman, 2010). This work offers the computational counterpart to the strong embodiment claim of Gallese and Lakoff (2005), but issues of how brain structures may support the metaphorical mappings and the constructions (and their assemblage) are left unaddressed.

Where ECG is based on the study of motor schemas, Arbib and Lee (2008) "lifted" the classic VISIONS model of the role of visual schemas in scene perception (Hanson & Riseman, 1978) to package an abstraction from a visual schema assemblage into a semantic representation (SemRep) that could be made accessible to the language system. They developed *Template Construction Grammar* (TCG) to show how competition and cooperation between constructions, represented as schemas, was able to map the SemRep onto utterances through a dynamic and incremental process of cooperative computation. Barrès and Lee (2013) made a first step towards extending TCG to comprehension, arguing that behavioral results of agrammatic patients performing sentence–picture matching tasks highlight the necessity of two functionally separated routes that may enter into cooperative computation with each other to generate an associated semantic representation that can in turn interact with visual schemas. One route ignores grammatical information and generates semantic representations on the basis of world knowledge alone, while the other uses the grammatical knowledge to dynamically map form onto semantics. This grammatical route is not "purely syntactic" but also incorporates the role of semantic constraints associated with constructions such as those required to explain why one can say "He poured water in the glass" and "He filled the glass with water" but not "He poured the glass with water" or "He filled water in the glass". Incorporating such effects into a computational model will be necessary to account for studies reporting that patients can be specifically impaired in processing such construction-based semantic constraints while retaining access to conceptual knowledge of the relevant actions (Kemmerer, 2000).

A further source of inspiration comes from a branch of robotics that asks how (artificial) agents that perceive and act in the physical world can establish and use a language to communicate as a result of, e.g., evolutionary language games repeatedly played within a community of embodied robotic agents. The Talking Heads experiment (Steels, 1999) consisted in a naming game: at each turn, two robots, selected from a population of agents, are placed in front of a visual scene composed of colored geometrical figures. One (the speaker) picks a figure in the scene (a topic) and orients its camera towards it. Then it tries to communicate to the other robot (the hearer) what it selected by producing words from its lexicon. The other robot is endowed with the capacity to use the sensor orientation of the speaker to orient its attention towards the generally relevant area of the visual scene. Upon hearing the words produced by the speaker, the hearer has to guess what the figure is and "point" towards it by orienting its camera. If the hearer is wrong, the speaker then points to the correct figure. Given the proper learning rules, it was shown that, starting from random idiosyncratic lexicons for each agent, a shared lexicon could self-organize and stabilize in the population. Parity of meaning is therefore achieved as an emergent property of embodied language use that results in the alignment of cognitive content. The embodied nature of the agents plays a central role in this process. *But* this is the embodiment of shared attention; it is not embodiment related to the meaning of the concept. And recall our discussion in Section 3.3. The fact that the acquisition of the name for a concept requires embodiment does *not* imply that later use of the concept is embodied. There need be nothing action-oriented or embodied when recognizing, e.g., a blue triangle save in specific contexts, such as preparing to draw a copy, or inserting a puzzle piece in the appropriate slot.

In further work, vertical transmission across generations of agents was added and the linguistic representations were expanded from lexical items to constructions grounded in sensory-motor representation using *Fluid Construction Grammar*, FCG (Steels, 2004). Beuls and Steels (2013) were able, using further evolutionary games, to show the emergence, evolution, and cultural transmission of grammatical agreement in a population through repeated linguistic interactions between embodied agents. We are here in the realm of historical linguistics, not in the realm of the biological evolution of the language-ready brain. An open challenge is to understand in what way the assumptions built into use of FCG can be used to postulate features that must have evolved in the human language-ready brain. Of course the same applies to ECG and TCG, and much effort will be needed to reconcile the different assumptions built into these three frameworks.

Going back to the question of gesture acquisition in apes developed in Section 3.2, the question that Arbib et al. (2014) asked is not simply how dyadic praxic interactions between agents can lead, by a process of ontogenetic ritualization, to the emergence of meaningful gestures. We also sought to computationally understand the neural changes necessary to account for the fact that this process can take place during interaction between mother and child for chimpanzees but not for macaques. There is one approach to construction grammar that does take neural data on macaque and human into account. The starting point is a model (Dominey, Arbib, & Joseph, 1995) showing how association of sequences of saccadic eye movements with appropriate cues can be performed by a recurrent prefrontal cortical network that encodes sequential structure coupled with cortico-striatal connections that learn to associate different prefrontal states with appropriate actions. Turning to a neurolinguistic model that employs Construction Grammar, Dominey, Hoen, Blanc, and Lelekov-Boissard (2003) distinguish serial, temporal, and abstract structure in sequences. Serial structure learning involves the problem of how many elements must be remembered in order to correctly predict the next element in a series. Temporal structure can be thought of as "rhythm", i.e. the duration of elements in a sequence and pauses between them. Abstract structure is represented by generative rules describing the hierarchical organization of a sequence. Dominey, Hoen, and Inui (2006) consider constructions as templates consisting of a string of function words as well as slots into which a variety of open class elements (nouns, verbs, etc.) can be inserted in order to express novel meanings. Their simplifying assumption is that the string of function words uniquely indexes each construction. However, the constructions for *The boy threw Mary the bone* and *The brown dog licked the bone* both have the form ⟨*the* _ _ _ *the* _⟩ but correspond to different constructions. For Dominey et al., determining the construction type for a given sentence consists in analyzing the sentence as a sequence of open class and closed class elements, and then performing sequence reorganization on this sequence using structures inspired by the earlier model of the role of basal ganglia in learning associations and sequences. The system uses categorization of the construction to assign content words in a simple sentence to the semantic roles ACTION, AGENT, OBJECT and

Please cite this article as: Arbib, M. A., et al. Language is handy but is it embodied? Neuropsychologia (2013), http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004

RECIPIENT of the event specified by the input sentence. Dominey, Hoen, and Inui hypothesized that insertion of the referent semantic content is realized in pars triangularis BA45 and that the mapping from form to meaning take place in frontal cortical regions including BA 44, 46 & 6. Comparison of fMRI brain activation in sentence processing and nonlinguistic sequence mapping tasks (Hoen, Pachot-Clouard, Segebarth, & Dominey, 2006) found that a common cortical network including BA44 was involved in both the processing of sentences and abstract structure in non-linguistic sequences whereas BA45 was exclusively activated in sentence processing. The 2006 model did not really address "going hierarchical", though it did offer a simple account of the recognition of relative clauses. In any case, this work does establish a basis within modeling for a testable hypothesis which distinguishes the processing of sentences (a human specialty) from mechanisms for processing abstract structure in non-linguistic sequences (which may be shared with macaques).

5. Conclusion

The *core* notions of object and action as encountered in everyday embodied behavior are universal, but the conceptualized structure of space and its domains of interaction can vary drastically between species, and a human's notion of space may range from embodied experience to the furthest abstractions of mathematics and physics. We have examined Marc Jeannerod's view of perception and cognition *grounded* in the service of action while stressing that evolution builds atop species-specific schemas (computational neuroethology) to support schemas of increasing abstraction (cognitive neuroscience). This trend in humans is the basis for, and is accelerated by, the evolution of language. We offered an approach to a *graded* notion of embodiment. This approach is not an end in itself but rather the beginning of renewed efforts to develop a computational neuroscience of language which offers a role for embodiment in some instances of language processing, but does not preclude the "escape" from embodiment. In some contexts, sensorimotor systems may participate in mediating the perception and decision processes in language use (as distinct from the embodiment of the perception of auditory form and production of articulatory form), but in other contexts, the mediation of these processes would depend very little on sensorimotor systems.

For us, the matter of grading embodiment is less important than the challenge of extending and (where necessary) correcting the Mirror System Hypothesis in concert with a new *comparative neuro-primatology* linking empirical analysis to increasingly powerful computational brain models (with increasing emphasis on interacting dyads). Such models will include, but not be limited to, those situations (briefly reviewed in Section 3.5) which do indeed engage "symptoms of embodiment". For example, Arbib, Bonaiuto, Jacobs, and Frey (2009) outlined strategies for extending our previous computational models of macaque neural circuitry for generating and recognizing actions (Section 1.3) to address data on tool use from macaque neurophysiology (Umiltà et al., 2008) and human brain imaging (Jacobs, Danielmeier, & Frey, 2009); while Section 3.2 showed how such models might be extended to chart neural mechanisms underlying gesture acquisition in apes.

Turning to the human, much work is needed to understand how the human brain is uniquely able to make the transition from simple to complex imitation and to lift that capability from praxic to communicative action, to form increasingly abstract concepts, to then form words and link them to concepts of increasing abstraction and assemble them through constructions of varied levels of generality, and to do so in a way that allows new metaphors to form to repeatedly extend the communicative range open to most members of a particular language community.

Acknowledgment

This material is based in part on work supported by the National Science Foundation under Grant no. BCS-1343544 (Michael A. Arbib, Principal Investigator).

References

Aravena, P., Delevoye-Turrell, Y., Deprez, V., Cheylus, A., Paulignan, Y., Frak, V., et al. (2012). Grip force reveals the context sensitivity of language-induced motor activity during "action words" processing: Evidence from sentential negation. *PLoS One*, *7*, e50287.

Arbib, M. A. (1972). *The metaphorical brain: An introduction to cybernetics as artificial intelligence and brain theory*. New York: Wiley-Interscience.

Arbib, M. A. (1981). Perceptual structures and distributed motor control. In: V. B. Brooks (Ed.), *Handbook of physiology — The nervous system II. Motor control* (pp. 1449–1480). Bethesda, MD: American Physiological Society.

Arbib, M. A. (1987). Levels of modelling of visually guided behavior (with peer commentary and author's response). *Behavioral and Brain Sciences*, *10*, 407–465.

Arbib, M. A. (2002). The mirror system, imitation, and the evolution of language. In: K. Dautenhahn, & C. L. Nehaniv (Eds.), *Imitation in animals and artifacts. Complex adaptive systems* (pp. 229–280). Cambridge, MA: The MIT Press.

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. (2012). *How the brain got language: The mirror system hypothesis*. New York & Oxford: Oxford University Press.

Arbib, M. A., Bonaiuto, J. J., Jacobs, S., & Frey, S. H. (2009). Tool use and the distalization of the end-effector. *Psychology Research*, *73*, 441–462.

Arbib, M. A., & Caplan, D. (1979). Neurolinguistics must be computational. *Behavioral and Brain Sciences*, *2*, 449–483.

Arbib, M. A., Ghanesh, V., & Gasser, B. (2013). Dyadic brain modeling, ontogenetic ritualization of gesture in apes, and the contributions of primate mirror neuron systems. *Philosophical Transactions of the Royal Society B* (in press)

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

Arbib, M. A., Iberall, T., & Lyons, D. (1985). Coordinated control programs for control of the hands. In: A. W. Goodwin, & I. Darian-Smith (Eds.), *Hand function and the neocortex* (pp. 111–129). Berlin: Springer-Verlag.

Arbib, M. A., & Lee, J. Y. (2008). Describing visual scenes: Towards a neurolinguistics based on construction grammar. *Brain Research*, *1225*, 146–162.

Arévalo, A. L., Baldo, J. V., & Dronkers, N. F. (2012). What do brain lesions tell us about theories of embodied semantics and the human mirror neuron system? *Cortex*, *48*, 242–254.

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

Aziz-Zadeh, L., & Damasio, A. (2008). Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology*, *102*, 35–39.

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: CB*, *16*, 1818–1823.

Barrès, V., & Lee, J. Y. (2013). Template construction grammar: From visual scene description to language comprehension and agrammatism. *Neuroinformatics*, http://dx.doi.org/10.1007/s12021-12013-19197-y.

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*, 577–609 (discussion 610–660).

Bergen, B. K., & Chang, N. (2005). *Embodied construction grammar in simulation-based language understanding. Construction grammars: Cognitive grounding and theoretical extensions*. .

Beuls, K., & Steels, L. (2013). Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PloS One*, 8.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*, 527–536.

Bonaiuto, J. J., & Arbib, M. A. (2010). Extending the mirror neuron system model II: What did I just do? A new role for mirror neurons. *Biological Cybernetics*, *102*, 341–359.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., et al. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: An fMRI study. *Journal of Cognitive Neuroscience*, *16*, 114–126.

Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Science*, *21*, 667–684 (discussion 684–721).

Byrne, R. W., & Whiten, A. (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys. Apes and humans*. Oxford: Clarendon Press.

Call, J., & Tomasello, M. (2007). *The gestural communication of apes and monkeys*. New York: Lawrence Erlbaum Associates.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*

Please cite this article as: Arbib, M. A., et al. Language is handy but is it embodied? *Neuropsychologia* (2013), http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, *30*, 687–696.

Cheney, D. L., & Seyfarth, R. M. (2007). *Baboon metaphysics: The evolution of a social mind*. Chicago: University Of Chicago Press.

Clark, A. (2008). Pressing the flesh: A tension in the study of the embodied, embedded mind?*. *Philosophy and Phenomenological Research*, *76*, 37–59.

Corballis, M. C. (2002). *From hand to mouth, the origins of language*. Princeton, New Jersey: Princeton University Press.

Corina, D. P., Poizner, H., Bellugi, U., Feinberg, T., Dowd, D., & O'Grady-Batch, L. (1992). Dissociation between linguistic and nonlinguistic gestural systems: a case for compositionality. *Brain and Language*, *43*, 414–447.

Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge University Press.

D'Ausilio, A., Pulvermuller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*, 381–385.

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

de Bruin, L., & Gallagher, S. (2012). Embodied simulation, an unproductive explanation: Comment on Gallese and Sinigaglia. *Trends in Cognitive Sciences*, *16*, 98–99.

Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012). Identification of the social and cognitive processes underlying human cumulative culture. *Science*, *335*, 1114–1118.

Deaner, R. O., Khera, A. V., & Platt, M. L. (2005). Monkeys pay per view: Adaptive valuation of social images by rhesus macaques. *Current Biology*, *15*, 543–548.

Dessalles, J. -L. (2008). From metonymy to syntax in the communication of events. *Interaction Studies*, *9*, 51–66.

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

Didday, R. L. (1970). *The simulation and modelling of distributed information processing in the frog visual system (Ph.D. thesis)*. Palo Alto, California: Stanford University (unpublished).

Diessel, H. (2013). Where does language come from? Some reflections on the role of deictic gesture and demonstratives in the evolution of language. *Language and Cognition*, *5*, 239–249.

Dominey, P. F., Arbib, M. A., & Joseph, J.-P. (1995). A model of corticostriatal plasticity for learning oculomotor associations and sequences. *Journal of Cognitive Neuroscience*, *7*, 311–336.

Dominey, P. F., Hoen, M., Blanc, J.-M., & Lelekov-Boissard, T. (2003). Neurological basis of language and sequential cognition: Evidence from simulation, aphasia, and ERP studies. *Brain and Language*, *86*, 207–225.

Dominey, P. F., Hoen, M., & Inui, T. (2006). A neurolinguistic model of grammatical construction processing. *Journal of Cognitive Neuroscience*, *18*, 2088–2107.

Dove, G. (2010). On the need for embodied and dis-embodied cognition. *Frontiers in Psychology*, *1* (242–242).

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *6*, 178–190.

Engel, A. K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences*, *17*, 202–209.

Ewert, J.-P., & Arbib, M. A. (1989). *Visuomotor coordination: Amphibians, comparisons, models and robots*. New York: Plenum Press.

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

Feldman, J. (2010). Embodied language, best-fit analysis, and formal compositionality. *Physics of Life Reviews*, *7*, 385–410.

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

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.

Fujii, N., Hihara, S., & Iriki, A. (2008). Social cognition in premotor and parietal cortex. *Social Neuroscience*, *3*, 250–260.

Fujii, N., Hihara, S., Nagasaka, Y., & Iriki, A. (2008). Social state representation in prefrontal cortex. *Social Neuroscience*, *4*, 73–84.

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

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *TRENDS in Cognitive Sciences*, *8*, 396–403.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*, 455–479.

Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences*, *15*, 512–519.

Garagnani, M., Wennekers, T., & Pulvermüller, F. (2008). A neuroanatomically-grounded Hebbian learning model of attention-language interactions in the human brain. *European Journal of Neuroscience*, *27*, 492–513.

Gasser, B., Cartmill, E., & Arbib, M. A. (2013). Ontogenetic ritualization of primate gesture as a case study in dyadic brain modeling. *Neuroinformatics*, http://dx.doi.org/10.1007/s12021-12013-19182-12025.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Goldman, A. (2012). A moderate approach to embodied cognitive science. *Review of Philosophy and Psychology*, *3*, 71–88.

Goldman, A., & de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences*, *13*, 154–159.

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, *349*, 154–156.

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*, 377–388.

Grice, H. P. (1969). Utterer's meaning and intention. *The Philosophical Review*, *78*, 147–177.

Grice, H. P. (1975). *Logic and conversation*, 1975.

Grossman, M., Anderson, C., Khan, A., Avants, B., Elman, L., & McCluskey, L. (2008). Impaired action knowledge in amyotrophic lateral sclerosis. *Neurology*, *71*, 1396–1401.

Hanson, A. R., & Riseman, E. M. (1978). VISIONS: A computer system for interpreting scenes. In: A. R. Hanson, & E. M. Riseman (Eds.), *Computer vision systems* (pp. 129–163). New York: Academic Press.

Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (2005). *The World Atlas of language structures*. Oxford: Oxford University Press.

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

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The language faculty: What is it, who has it, and how did it evolve? *Science*, *298*, 1568–1579.

Hecht, E. E., Gutman, D. A., Preuss, T. M., Sanchez, M. M., Parr, L. A., & Rilling, J. K. (2013). Process versus product in social learning: comparative diffusion tensor imaging of neural systems for action execution–observation matching in macaques, chimpanzees, and humans. *Cerebral Cortex*, *23*, 1014–1024.

Hewes, G. W. (1973). Primate communication and the gestural origin of language. *Current Anthropology*, *12*, 5–24.

Hobaiter, C., & Byrne, R. W. (2011). The gestural repertoire of the wild chimpanzee. *Animal Cognition*, *14*, 745–767.

Hoen, M., Pachot-Clouard, M., Segebarth, C., & Dominey, P. F. (2006). When Broca experiences the Janus syndrome: An ER-fMRI study comparing sentence comprehension and cognitive sequence processing. *Cortex*, *42*, 605–623.

Hoff, B., & Arbib, M. A. (1991). A model of the effects of speed, accuracy and perturbation on visually guided reaching. In: R. Caminiti, P. B. Johnson, & Y. Burnod (Eds.), *Control of arm movement in space: Neurophysiological and computational approaches* (pp. 285–306). Heidelberg, New York: Springer-Verlag.

Hoff, B., & Arbib, M. A. (1993). Models of trajectory formation and temporal interaction of reach and grasp. *Journal of Motor Behavior*, *25*, 175–192.

Humphrey, N. K. (1970). What the frog's eye tells the monkey's brain. *Brain Behavior and Evolution*, *3*, 324–337.

Iberall, T., Bingham, G., & Arbib, M. A. (1986). Opposition space as a structuring concept for the analysis of skilled hand movements. In: H. Heuer, & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 158–173). Berlin: Springer-Verlag.

Ingle, D. J. (1968). Visual releasers of prey catching behaviour in frogs and toads. *Brain, Behavior and Evolution*, *1*, 500–518.

Ingle, D. J., Schneider, G. E., Trevarthen, C. B., & Held, R. (1967). Locating and identifying: Two modes of visual processing (a symposium). *Psycholische Forschung*, 31.

Jacob, P. (2012). Embodying the mind by extending it. *Review of Philosophy and Psychology*, *3*, 33–51.

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: A critique. *Trends in Cognitive Sciences*, *9*, 21–25.

Jacobs, S., Danielmeier, C., & Frey, S. H. (2009). Human anterior intraparietal and ventral premotor cortices support representations of grasping with the hand or a novel tool. *Journal of Cognitive Neuroscience*, *22*, 2594–2608.

Jeannerod, M. (1994). The representing brain. Neural correlates of motor intention and imagery. *Behavioral and Brain Science*, *17*, 187–245.

Jeannerod, M. (1997). *The cognitive neuroscience of action*. Oxford: Blackwell.

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage*, *14*, S103–S109.

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

Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: The cortical mechanisms of visuomotor transformation. *Trends in Neuroscience*, *18*, 314–320.

Jeannerod, M., & Biguer, B. (1982). Visuomotor mechanisms in reaching within extra-personal space. In: D. J. Ingle, R. J.W. Mansfield, & M. A. Goodale (Eds.), *Advances in the analysis of visual behavior* (pp. 387–409). Cambridge, MA: The MIT Press.

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

Jeannerod, M., Michel, F., & Prablanc, C. (1984). The control of hand movements in a case of hemianaesthesia following a parietal lesion. *Brain*, *107*, 899–920.

Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Lüdtke, J. (2007). Experiential simulations of negated text information. *The Quarterly Journal of Experimental Psychology*, *60*, 976–990.

Kemmerer, D. (2000). Grammatically relevant and grammatically irrelevant features of verb meaning can be independently impaired. *Aphasiology*, *14*, 997–1020.

Kemmerer, D., Rudrauf, D., Manzel, K., & Tranel, D. (2012). Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *48*, 826–848.

Kimura, D. (1993). *Neuromotor mechanisms in human communication*. New York, Oxford: Oxford University Press.

Kita, S., & Davies, T. S. (2009). Competing conceptual representations trigger co-speech representational gestures. *Language and Cognitive Processes*, *24*, 761–775.

Laland, K. N., Odling-Smee, F. J., & Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences*, *23*, 131–146.

Langacker, R. W. (1986). *Foundations of cognitive grammar*. Vol. 1. Stanford: Stanford University Press.

Lesser, V. R., Fennel, R. D., Erman, L. D., & Reddy, D. R. (1975). Organization of the HEARSAY-II speech understanding system. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *23*, 11–24.

Lettvin, J. Y., Maturana, H., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog brain. *Proceedings of the IRE*, *47*, 1940–1951.

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

Lidz, J., Pietroski, P., Halberda, J., & Hunter, T. (2011). Interface transparency and the psychosemantics of most. *Natural Language Semantics*, *19*, 227–256.

MacNeilage, P. F., & Davis, B. L. (2005). The frame/content theory of evolution of speech: A comparison with a gestural-origins alternative. *Interaction Studies*, *6*, 173–199.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology*, *102*, 59–70.

Marghetis, T., & Núñez, R. (2013). The motion behind the symbols: A vital role for dynamism in the conceptualization of limits and continuity in expert mathematics. *Topics in Cognitive Science*, *5*, 299–316.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2010). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*

Mithen, S. (2005). *The singing Neanderthals: The origins of music, language, mind & body*. London: Weidenfeld & Nicholson.

Moulin-Frier, C., & Arbib, M. A. (2013). Recognizing speech in a novel accent: The motor theory of speech perception reframed. *Biological Cybernetics*, *107*, 421–447, http://dx.doi.org/10.1007/s00422-00013-00557-00423.

Narayanan, S. (1999). *Moving right along: A computational model of metaphoric reasoning about events*. Ph.D. 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*, 116–140.

Papeo, L., Negri, G. A. L., Zadini, A., & Rumiati, R. I. (2010). Action performance and action-word understanding: Evidence of double dissociations in left-damaged patients. *Cognitive Neuropsychology*, *27*, 428–461.

Paulignan, Y., Jeannerod, M., MacKenzie, C., & Marteniuk, R. (1991). Selective perturbation of visual input during prehension movements. 2. The effects of changing object size. *Experimental Brain Research*, *87*, 407–420.

Paulignan, Y., MacKenzie, C., Marteniuk, R., & Jeannerod, M. (1991). Selective perturbation of visual input during prehension movements. 1. The effects of changing object position. *Experimental Brain Research*, *83*, 502–512.

Perlman, M., Tanner, J. E., & King, B. J. (2012). In: S. Pika, & K. Liebal (Eds.), *A mother gorilla's variable use of touch to guide her infant: Insights into iconicity and the relationship between gesture and action* (1 ed.). Amsterdam: John Benjamins Publishing Company.

Piaget, J. (1954). *The construction of reality in the child*. New York: Norton.

Pickering, M. J., & Garrod, S. (2013a). Forward models and their implications for production, comprehension, and dialogue. *Behavioral and Brain Sciences*, *36*, 377–392.

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

Poizner, H., Klima, E. S., & Bellugi, U. (1987). *What the hands reveal about the brain*. Cambridge, MA: MIT Press.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*, 576–582.

Pulvermüller, F. (2013). Semantic embodiment, disembodiment or misembodiment? In search of meaning in modules and neuron circuits. *Brain and Language*, *127*, 86–103.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*, 351–360.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*, 793–797.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., et al. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience*, *11*, 426–428.

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

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Experimental Brain Research*, *71*, 491–507.

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

Santos, G. S., Nagasaka, Y., Fujii, N., & Nakahara, H. (2011). Encoding of social state information by neuronal activities in the macaque caudate nucleus. *Social Neuroscience*, *7*, 42–58.

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

Sperber, D., & Wilson, D. (1986/95). *Relevance: Communication and cognition*. Oxford: Blackwell.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, *17*, 3–23.

Steels, L. (1999). *The talking heads experiment. Words and meaning (Special preedition)*, Vol. I. Brussels: Vrije Universiteit Brussel.

Steels, L. (2004). *Fluid construction grammars. A Brief Tutorial*. .

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, *30*, 299–351.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, *12*, 49–100.

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E. m., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current Anthropology*, *53*, 673–692.

Tomasino, B., Weiss, P. H., & Fink, G. R. (2010). To move or not to move: Imperatives modulate action-related verb processing in the motor system. *Neuroscience*, *169*, 246–258.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, *114*, 285–310.

Umiltà, M. A., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., et al. (2008). When pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences of United States of America*, *105*, 2209–2213.

Whiten, A., Goodall, J., McGrew, C., Nishida, T., Reynolds, V., Sugiyama, Y., et al. (1999). Cultures in chimpanzees. *Nature*, *399*, 682–685.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625–636.

Wittgenstein, L. (1921). *Tractatus logico-philosophicus*. London: Routledge & Kegan Paul.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*, 593–602.

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*, 1317–1329.

Please cite this article as: Arbib, M. A., et al. Language is handy but is it embodied? *Neuropsychologia* (2013), http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.004