is saying, "and he goes *up through* the pipe this time," with the gesture occurring during the boldfaced portion (the illustration captures the moment when the speaker says the vowel of "through"). Coexpressively with "up," her hand rose upward, and coexpressively with "through," her fingers spread outward to create an interior space. These took place together and were synchronized with "up through," the linguistic package that combines the same meanings.

The effect is a uniquely gestural way of packaging meaning – something like "rising hollowness," which does not exist as a semantic package of English at all. Speech and gesture, at the moment of their synchronization, were coexpressive. The very fact there is shared reference to the character's climbing up inside the pipe makes clear that it is being represented by the speaker in two ways simultaneously – analytic/combinatoric in speech and global/synthetic in gesture. We suggest it was this very simultaneous combination of opposites that evolution seized upon.

2. When signs and speech do combine in contemporary human performance, they do not synchronize. Kendon (1988) observed sign languages employed by aboriginal Australian women – full languages developed culturally for (rather frequent) speech taboos – which they sometimes combine with speech. The relevant point is that in producing these combinations, speech and sign start out synchronously, but then, as the utterance proceeds, speech outruns the semantically equivalent signs. The speaker stops speaking until the signs catch up and then starts over, only for speech and signs to pull apart again. If, in the evolution of language, there had been a similar doubling up of signs and speech, as the supplanting scenario implies, they too would have been driven apart rather than into synchrony, and for this reason, too, we doubt the replacement hypothesis.

3. The Wundt/Saussure "double essence" of gesture and language appears to be carried by a dedicated thought-hand-language circuit in the brain. This circuit strikes us as a prime candidate for an evolutionary selection at the foundation of language. It implies that the aforementioned *combinations* of speech and gesture were the selected units, not gesture first with speech supplanting or later joining it. We observe this circuit in the unique neurological case of I.W., who lost all proprioception and spatial position sense from the neck down at age 19, and has since taught himself to move using vision and cognition. The thought-language-hand link, located presumably in Broca's area, ties together language and gesture, and, in I.W., survives and is partly dissociable from instrumental action.

We can address Arbib's pantomime model by observing the kinds of gestures the dedicated link sustains in I.W.'s performance, in the absence of vision: his gestures are (1) coexpressive and synchronous with speech; (2) not supplemental; and (3) not derivable from pantomime. I.W. is unable to perform instrumental actions without vision but continues to perform speech-synchronized, coexpressive gestures that are virtually indistinguishable from normal (topokinetic accuracy is reduced but morphokinetic accuracy is preserved) (Cole et al. 2002). His gestures without vision, moreover, minimize the one quality that could be derived from pantomime, a so-called "first-person" or "character" viewpoint, in which a gesture replicates an action of a character (cf. McNeill 1992).

More generally, an abundance of evidence demonstrates that spontaneous, speech-synchronized gestures should be counted as part of language (McNeill 1992). Gestures are frequent (accompanying up to 90% of utterances in narrations). They synchronize exactly with coexpressive speech segments, implying that gesture and related linguistic content are coactive in time and jointly convey what is newsworthy in context. Gesture adds cohesion, gluing together potentially temporally separated but thematically related segments of discourse. Speech and gesture develop jointly in children, and decline jointly after brain injury. In contrast to cultural emblems, such as the "O.K." sign, speech-synchronized gestures occur in all languages, so far as is known. Finally, gestures are not "signs" with an independent linguistic code. Gestures exist only in combination with speech, and are not themselves a coded system.

Arbib's gesture-first. Arbib's concept of an expanding spiral may avoid some of the problems of the supplanting mechanism. He speaks of scaffolding and spiral expansion, which appear to mean, in both cases, that one thing is preparing the ground for or propping up further developments of the other thing – speech to gesture, gesture to speech, and so on. This spiral, as now described, brings speech and gesture into temporal alignment (see Fig. 6 in the target article), but also implies two things juxtaposed rather than the evolution of a single "thing" with a double essence. Modification to produce a dialectic of speech and gesture, beyond scaffolding, does not seem impossible. However, the theory is still focused on gestures of the wrong kind for this dialectic – in terms of Kendon's Continuum (see McNeill 2000 for two versions), signs, emblems, and pantomime. Because it regards all gestures as simplified and meaning-poor, it is difficult to see how the expanding spiral can expand to include the remaining point on the Continuum, "gesticulations" – the kind of speech-synchronized coexpressive gesture illustrated above.

A compromise is that pantomime *was* the initial protolanguage but was replaced by speech *plus gesture*, leading to the thought-language-hand link that we have described. This hypothesis has the interesting implication that different evolutionary trajectories landed at different points along Kendon's Continuum. One path led to pantomime, another to coexpressive and speech-synchronized gesticulation, and so on. These different evolutions are reflected today in distinct ways of combining movements with speech. Although we do not question the importance of extending the mirror system hypothesis, we have concerns about a theory that predicts, as far as gesture goes, the evolution of what did not evolve instead of what did.

Meaning and motor actions: Artificial life and behavioral evidence

Domenico Parisi,[a] Anna M. Borghi,[b] Andrea Di Ferdinando,[c] and Giorgio Tsiotas[b]

[a]*Institute of Cognitive Science and Technologies, National Research Council, Rome 00137, Italy;* [b]*Department of Psychology, University of Bologna, Bologna 40127, Italy;* [c]*Department of Psychology, University of Padua, Padua 35131, Italy.* **d.parisi@istc.cnr.it**
http://gral.ip.rm.cnr.it/dparisi/ **annamaria.borghi@unibo.it**
http://gral.ip.rm.cnr.it/borghi/ **andrea.diferdinando@unipd.it**
http://ccnl.psy.unipd.it/di_ferdinando.html **giorgio@tsiotas.com**

Abstract: Mirror neurons may play a role in representing not only signs but also their meaning. Because actions are the only aspect of behavior that are inter-individually accessible, interpreting meanings in terms of actions might explain how meanings can be shared. Behavioral evidence and artificial life simulations suggest that seeing objects or processing words referring to objects automatically activates motor actions.

Arbib argues that the vocal signs of human language are probably evolved from the gestural signs of some protolanguage, and this might explain why the production of vocal signs in the human brain is controlled by Broca's area – which corresponds to area V5 in monkeys' brain – which controls manual actions. The discovery of neurons in both areas that are activated both when a manual action is executed and when it is observed in others (mirror neurons) reinforces this interpretation, because language is based on what Arbib calls the parity requirement, according to which what counts for the speaker must count approximately the same for the hearer.

However, language is not only signs but is signs plus the meaning of signs. Mirror neurons tend to be invoked to explain the production of linguistic signs but they may also play an important role in the representation of the meaning of those signs. If meanings are interpreted as categories of entities in the environment, one can argue that these categories are represented in the brain in

https://doi.org/10.1017/S0140525X05370038 Published online by Cambridge University Press

terms of the motor actions that we execute on them. Two entities are included in the same category, that is, they evoke the same pattern of neural activity, if we tend to execute the same actions on them, whereas two entities are included in different categories if we tend to execute different actions on them.

If we interpret not only signs but also their meaning in terms of motor actions, we can understand how meanings can be shared between speakers and hearers. Motor actions are the only aspect of behavior which is inter-individually accessible. A has no direct access to what B perceives, feels, or thinks, but only to how B moves its body and to the consequences of these movements. Meanings can be shared if the categories they refer to are represented in terms of motor actions in the brain of both speaker and hearer. Mirror neurons can play a role not only with respect to the motor (phono-articulatory) actions that result in the production of vocal signs but also with respect to the motor actions of all kinds that we execute on the entities that vocal signs refer to. A gestural origin of human language may have facilitated the emergence of shared meanings. As Arbib recognizes, gestural signs are more iconic than vocal signs, which means that gestural signs are motor actions which physically resemble the motor actions that we execute on the entities they refer to. Vocal signs are arbitrary, that is, non-iconic, but they may have exploited the already existing shared meanings neurally represented in terms of inter-interindividually accessible motor actions executed on objects.

Artificial life simulations and experiments suggest that seeing objects or processing words referring to objects automatically activates canonical actions we perform on them, particularly reaching and grasping movements. Borghi and colleagues (Borghi et al. 2002, 2005, submitted; Di Ferdinando & Parisi 2004) evolved simulated organisms using a genetic algorithm (Holland 1992). Each organism lives in a bidimensional environment containing four objects, either upright or reversed, with a handle protruding on the right or on the left. The organism possesses a visual system allowing it to see different objects, one at a time, and a motor system consisting of a single arm composed of two movable segments; the arm sends proprioceptive information to the organism, specifying the arm's current position. The organism's nervous system was simulated with a neural network (Fig. 1).

Organisms learned to reach the handle of the object independently of its position (Task 1) and then they learned to reach one of two buttons located below the handle to decide whether the object was upright or reversed (Task 2). In one condition, the button to be reached was on the same side of the object's handle; in another condition, it was on the opposite side. Task 1 reproduced real-life experience; Task 2 replicated an experiment made by Tucker and Ellis (1998). When the handle location and the button were spatially compatible, learning occurred earlier (in terms of generations required to reach an optimal performance) than when they were not. The results suggest that affordances of objects become learned in association with successful reaching movements. Once reaching becomes established, seeing the handle of objects activates appropriate movements.

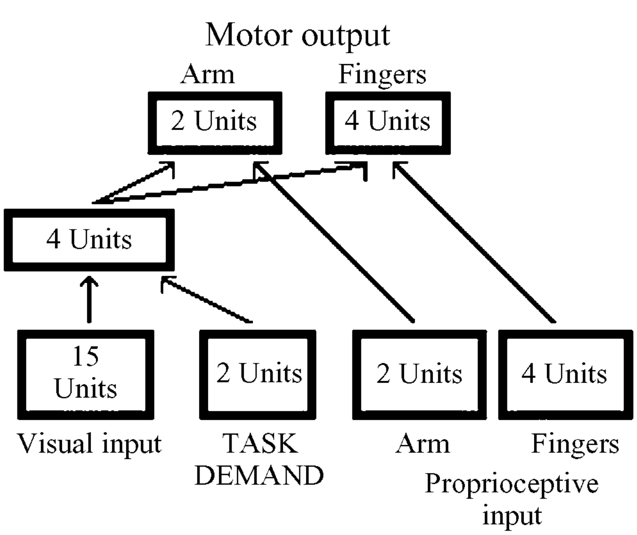

Figure 1 (Parisi et al.). Neural network controlling an arm.

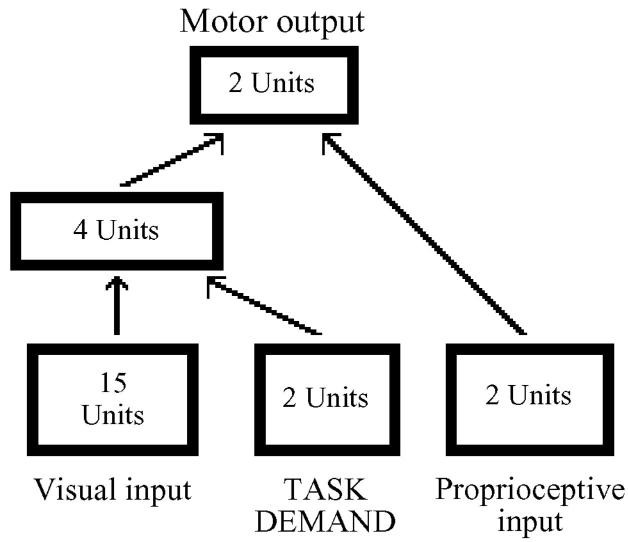

Figure 2 (Parisi et al.). Neural network controlling an arm and a hand.

Tsiotas et al. (in press) simulated an organism with a nervous system (Fig. 2) and with an arm terminating with a hand composed of two fingers, a thumb and an index, each composed by two segments.

The organism lived in a bidimensional environment containing four objects, either large or small and red or blue. First, the organism had to learn to grasp small objects with a precision grip and large objects with a power grip, then to decide the objects' color by grasping a small or a large button. Learning occurred earlier when the grip required to respond to the object and to decide the color was the same than when it was not, even if object size was irrelevant for the task (Ellis & Tucker 2000).

The inter-accessibility of these simple gestures which are automatically activated by objects, may have played a relevant role for language evolution. Crucially, these gestures are automatically activated not only by visual stimuli but by words, too (Gentilucci 2003b; Tucker & Ellis 2004). Borghi et al. (2004) found in a part-verification task that responding by moving the arm in a direction incompatible with the part location (e.g., responding downward to verify that a car has a roof, upward to verify that a car has wheels) was slow relative to responding in a direction compatible with the part location.

The presence of action-based compatibility effects also with words, argues for the involvement not only of the dorsal but also of the ventral system and of long-term knowledge in generating affordances: accordingly, these effects would be accounted for by long-term visuomotor associations between objects and actions executed on them.