The mirror-neuron system and its role in imitation and language.

Giacomo Rizzolatti and Giovanni Buccino

Dipartimento di Neuroscienze, Sezione di Fisiologia , Università di Parma, 39 Via Volturno, 43100 Parma, Italy.

Phone: +39 0521 903879; Fax:+39 0521 903900

Introduction

The human species is characterized by two fundamental cognitive abilities that are poorly developed or lack altogether in other primates: imitation learning and the faculty of language. These abilities are, obviously, not the only ones that differentiate humans from other primates including apes, but certainly are among the most important.

Until recently it was hard to indicate, even in very speculative terms, the neural mechanisms underlying imitation and the faculty of language. This changed with the discovery of mirror neurons. The functional properties of these neurons suggest that primates are endowed with a mechanism that directly maps an observed action onto its motor counterpart. This matching mechanism appears to be able to provide a solution to the basic problem of imitation, i.e. how an action described in visual terms may be replicated by the motor system, using completely different physiological parameters. The same mechanism may also give an account, although, obviously, highly speculative on the riddle of language evolution.

The organization of the present chapter is the following. We will first summarize the functional properties of mirror neurons, the basic neural elements that constitute the mirror-neuron system. We will compare then the properties of human mirror-neuron system with that of monkeys, examining in particular those properties that may account for the faculty of imitation in humans. In the last part of

the study we will examine the relations between mirror-neuron system and language.

Functional properties of F5 mirror neurons

Mirror neurons were originally discovered in the rostral part of the ventral premotor cortex of the macaque monkey (area F5). Like all neurons of this area, mirror neurons have motor properties. They code mostly distal hand actions such as grasping, holding, tearing, manipulating. Their defining functional characteristic is that they become active not only when the monkey does a particular action (like grasping an object) but also when it observes another individual (monkey or human) making a similar action. Mirror neurons do not respond to the sight of a hand mimicking an action or to meaningless intransitive movements. Similarly, they do not respond to the observation of an object alone, even when it is of interest to the monkey (Gallese et al. 1996; Rizzolatti et al. 1996.).

The vast majority of F5 mirror neurons show a marked similarity between the action effective in triggering them when observed and the action effective in activating them when executed. This sensory-motor congruence is occasionally extremely strict. In these cases the effective motor action and the effective observed action coincide both in terms of goal (e.g. grasping) and in terms of how the goal is achieved (e.g. precision grip). For most mirror neurons, however, the congruence is broader and is confined to the goal of the action.

Early studies on mirror neurons concerned essentially the upper sector of F5 where hand actions are mostly represented. Recently, a study was carried out on the properties of neurons located in the lower part of F5 where neuron activity is mostly related to mouth actions (Ferrari et al. 2003).

The results showed that about 25% of "mouth" neurons have mirror properties. According to the visual stimuli effective in triggering the neurons, two classes of mouth mirror neurons were distinguished: ingestive and communicative mirror neurons. Ingestive mirror neurons (80% of the recorded mouth mirror neurons) respond to the observation of actions related to ingestive functions (e.g. grasping food with the mouth). Virtually, all of them show a good correspondence between the effective observed and the effective executed action.

More intriguing are the properties of the communicative mirror neurons. The most effective observed action is for them a communicative gesture such as, for example, lip smacking. However, as the ingestive mirror neurons, they strongly discharge when the monkey actively performs an ingestive action (Figure 1).

This discrepancy between the effective visual input (communicative) and the effective active action (ingestive) is rather intriguing. Yet, there is evidence suggesting that, in evolution, communicative gestures, or at least some of them, derived from ingestive actions (see below). From this perspective one may argue that the communicative mouth mirror neurons found in F5 reflects a process of corticalization of communicative functions, not yet freed from their original ingestive basis.

An issue recently addressed was whether mirror neurons are able to recognize actions from their sound. Kohler et al. (2002) recorded F5 mirror neuron activity while the monkey was observing a "noisy" action (e.g. ripping a piece of paper), or was presented with the sound of the action without seeing it. The results showed that about 15% of mirror neurons responsive to presentation of actions accompanied by sounds also responded to the presentation of the sound alone. Most of them discharged specifically to the sound typical of the observed action. These neurons were dubbed "audio-visual" mirror neurons. The properties of audio-visual neurons strongly suggest that mirror-neuron system is involved in action recognition, whatever the modality through which the action is presented.

Cortical representation of action observation

Mirror neurons are not present only in area F5. A further cortical area where mirror neurons have been described is area 7 of Brodmann (1909) or area PF of Von Economo (1929) (Fogassi et al. 1998, Gallese et al. 2002). This area occupies the rostral part of the inferior parietal lobule. It receives input from the cortex buried in the superior temporal sulcus (STS) and sends an important output to ventral premotor cortex including area F5.

PF neurons are functionally heterogeneous. Most of them (about 90%) respond to sensory stimuli including complex visual stimuli such as biological motion. About 50% of PF neurons have, in addition, motor properties discharging when the monkey

performs specific movements or actions (Hyvarinen 1982, Fogassi et al. 1998, Gallese et al. 2002). Some of these neurons have clear mirror properties (Gallese et al. 2002).

Another cortical region in which neurons respond to the observation of actions done by others is the cortex of the superior temporal sulcus (STS) (Perrett et al. 1989, 1990, Jellema et al. 2000, see Jellema et al. 2002). Movements effective in eliciting neuron responses in this region are walking, turning the head, bending the torso, and moving the arms. A small set of STS neurons discharges also during the observation of goal directed hand-movements (Perrett et al. 1990).

If one compares functional properties of STS with F5 and PF mirror neurons, the major difference is that STS neurons do not discharge in association with monkeys' actions. STS neurons do not appear, therefore endowed with motor properties.

In conclusion, the cortical mirror neuron circuit of the monkey is formed by two main regions: the rostral part of the inferior parietal lobule and the ventral premotor cortex. The functional significance of this system (mirror neuron-system) will be discussed in the next paragraphs.

Function of the mirror-neuron system in the monkey: action understanding.

There are two main hypotheses concerning the functional role of mirror neurons. The first is that mirror neuron activity underlies imitation (see Jeannerod 1994), the second is that they are at basis of action understanding (see Rizzolatti et al. 2001).

These two hypotheses are not mutually exclusive. However, because imitation is present, among primates, only in humans and (probably) in apes (see Galef 1988, Whiten & Ham 1992, Byrne 1995, Tomasello & Call 1997, Visalberghi & Fragaszy 2001), the evolutionary most ancient function of mirror neurons cannot be imitation. As it will be discussed later, imitation is a cognitive faculty which evolved later from the mirror system following a the acquisition of new matching properties by mirror neurons.

How do mirror neurons mediate understanding of actions done by others? The proposed mechanism is rather simple. Each time an individual sees an action made by another individual, neurons that represent that action are activated in the observer's premotor cortex. This, automatically activated motor representation corresponds to that that is spontaneously generated during active action and whose outcome is known to the acting individual. Thus, the mirror system is able to transform visual information into knowledge (see Rizzolatti et al. 2001).

The mirror-neuron system in humans

A mirror-neuron system also exists in humans. Evidence in this sense has been provided by a series of experiments carried out with various techniques such as EEG, MEG, TMS and brain imaging (see Rizzolatti et al. 2001).

These studies demonstrated that the human and monkey mirror-neuron systems share the capacity to match observed actions onto their motor representations. However they also showed some important differences.

First, in humans the observation of meaningful hand actions without an object (mimed actions) activates the mirror-neuron system (Buccino et al. 2001; Grèzes et al 2003). The presence of an object appears to be necessary to activate the mirror neurons in the monkey (see above).

Second, TMS experiments showed that that there is a facilitation of the motor evoked potentials (MEPs) recorded from the muscles of the observer corresponding to those used by the actor performing the action, also when an individual observes intransitive meaningless arm/hand gestures (Fadiga et al. 1995; Iacoboni et al. 1999; Maeda et al. 2002). Intransitive actions do not appear to be able to activate mirror neurons in the monkey.

Third, recent data showed that, during action observation, there is not only a facilitation of MEPs recorded from the corresponding muscles, but this facilitation follows the time-course of the observed action (Gangitano et al. 2001).

In conclusion, these properties indicate that the human mirror-neuron system is able to describe both the goal of an action and how the goal is achieved. The capacity to imitate not only the goal, but also the way in which the goal is achieved is considered a necessary prerequisite for distinguishing true imitation from emulation and other pseudo-imitative behaviors.

Imitation

Imitation is often thought of as a cognitively undemanding, rather elementary form of behavior. Recent works, across a variety of sciences, showed, however, that is

not true. There is clear evidence that imitation is a faculty particularly developed in humans, intrinsically linked to language and culture.

The involvement of the mirror-neuron system in imitation was recently demonstrated by a series of brain imaging studies. Using the fMRI technique, Iacoboni et al. (1999) scanned normal human volunteers while they lifted a finger in response to: a) the same action presented on a screen ("imitation"), b) to a symbolic cue, or c) to a spatial cue. The results showed that the activation was stronger during imitation than during the other motor conditions in the *pars opercularis* of the left inferior frontal gyrus (IFG), the right anterior parietal region, the right parietal operculum, and the right STS region (see also Iacoboni et al. 2001). Experiments by Koski et al. (2002) confirmed the importance of Broca's area, in particular when the action to be imitated had a specific goal. Grèzes et al. (2003) obtained similar results.

Nishitani & Hari (2000, 2002) performed two studies, using magneto-encephalography (MEG), in which they investigated imitation of grasping actions and of facial movements, respectively. The first study confirmed the importance of the left IFG (Broca's area) in imitation. In the second study (Nishitani & Hari 2002), the authors asked volunteers to observe still pictures of verbal and non-verbal (grimaces) lip forms, to imitate them immediately after having seen them, or to make similar lip forms spontaneously. During lip forms observation, cortical activation progressed from the occipital cortex to the superior temporal region, the inferior parietal lobule, IFG (Broca's area), and finally to the primary motor cortex. The activation sequence during imitation of both verbal and non-verbal lip forms was the same as during observation.

In spite of some minor discrepancies, these data clearly show that the basic circuit underlying imitation coincides with that active during action observation. They also indicate that, in the posterior part of IFG, a direct mapping of the observed action and its motor representation takes place.

The importance of the "pars opercularis" of IFG in imitation was recently further demonstrated by Heiser et al. (2003), using repetitive TMS, a technique that transiently disrupts the functions of the stimulated area. The task used in the study was, essentially, the same as that of the fMRI study by Iacoboni et al. (1999). The results showed that following stimulation of both left and right Broca's area, there was significant impairment in imitation of finger movements. The effect was absent when finger movements were done in response to spatial cues.

Imitation learning.

In the experiments reviewed above individuals were asked to repeat "on line" highly practiced actions made by another individual. A similar strategy was also used by Tanaka et al. (2002), who asked volunteers to imitate "on line" relatively complex hand or arm postures. In all these experiments the imitation consisted in matching the observed movements or action to motor models already present in the parietal lobe and the premotor areas and to produce them. No motor learning was involved.

Buccino et al. (2004) recently addressed the issue of which cortical areas become active when individuals are required not to simply repeat an action or a

posture present in their motor repertoire, but to produce, on the basis of action observation, a *new motor pattern.* The basic task was imitation, by naive participants, of guitar chords played by an expert guitarist. By using an event-related fMRI paradigm, cortical activations were mapped during the following events: a) observation of the chords made by the expert player b) pause (new motor pattern formation and consolidation), c) execution of the observed chords, and d) rest. In addition to the imitation condition, there were three control conditions: observation of the guitar chords made by the player without any subsequent motor request, observation of the chords followed by the execution of an actions not related to guitar chord execution (e.g. scratching the guitar neck), free execution of guitar chords.

The results showed that during the event observation in the imitation condition there was activation of a cortical network formed by the inferior parietal lobule and the dorsal part of PMv plus the *pars opercularis* of IFG. This network coincided with that active during the event observation without instruction to imitate and during the event observation in order not to imitate. The strength of the activation was, however, much stronger during imitation. During the event observation in the imitation condition, but not during the event observation without further motor action, there was, in addition, activation of the superior parietal lobule, anterior mesial areas plus a modest activation of the middle frontal gyrus.

The activation during the pause event in imitation condition involved the same basic circuit as in event observation, but with some important differences: increase of the superior parietal lobule activation, activation of PMd, and, most interestingly, a

dramatic increase in extension and strength of the middle frontal cortex activation (area 46) and of the areas of the anterior mesial wall. Finally, during the execution event, the activation concerned, not surprisingly, the sensorimotor cortex contralateral to the acting hand.

These data show that the nodal centers for new motor pattern formation coincide with the nodal mirror-neuron system centers. fMRI experiments cannot obviously give information on the mechanism underlying imitation, yet, it is plausible (see the neurophysiological sections) that, during learning of new motor patterns by imitation, the observed actions are decomposed into elementary motor acts that activate, by mirror mechanism, the corresponding motor representations in PF and in PMv and in the *pars opercularis* of IFG. Once these motor representations are activated, they are re-combined, to fit the observed model. This re-combination appears to occur inside the mirror-neuron circuit with area 46 playing a fundamental orchestrating role.

Mirror-neuron system and language

Some years ago Rizzolatti and Arbib (1998) proposed that the mirror matching mechanism represents the basic mechanism from which language evolved. This proposal was based on the consideration that mirror neurons create a direct link between the sender of a message and its receiver. Through them, therefore, observing and doing become manifestations of a single communicative (and, later, linguistic) faculty rather than two separate abilities.

Conceptually, the theory of Rizzolatti and Arbib belongs to theories that postulate that speech evolved, mostly, from gestural communication (see Armstrong et al. 1995; Corballis 2002). Its novelty consists in the fact that it indicates a neurophysiological mechanism that may create a common (parity requirement), non-arbitrary, link between the communicating individuals. This link can hardly be created by sounds. Sounds, by their nature, cannot generate the shared, non-arbitrary knowledge that can be achieved through the involvement of the motor system.

Humans mostly communicate by sounds. Thus, it seems almost natural to consider human speech as an evolutionary extension of the sound-based animal communication. In fact, human speech and animals' calls are different phenomena. First of all, the structures underlying speech and animals' calls in mammals are completely different. Animals' calls are mediated primarily by the cingulate cortex and by diencephalic and brain stem structures (see Juergens 2002). In contrast, human speech has its core substrate in the perisylvian areas, including area 44, a premotor area. Second, speech is not necessarily linked to an emotional behavior, whilst animals' calls are. Third, speech is mostly a person-to-person communication. In contrast animal calls are, typically, directed to "everybody", rather than to a specific individual. Fourth, but not least, speech is endowed with combinatorial properties that are absent in animal communication. It is recursive and virtually limitless with respect to its scope of expression.

The dichotomy between a communicative system as that of animals' calls and the one that eventually led to speech is nicely described by sir Richard Paget (1930). He writes: "... it may be imagined that, in the early stages of human development

mankind roared and grunted and sung, on the one hand, to express his emotions, and gesticulated and grimaced on the other to explain his ideas. In some cases he may have used both methods together, as when the dog makes the threatening gestures of his teeth, and energizes or phonates this gesture by the addition of a laryngeal growl".

There is evidence that, in humans, some types of emotional communication are based on a specific type of mirror-neuron mechanism (Carr et al. 2003, Krolak-Salmon et al. 2003; Wicker et al. 2004). One may argue, therefore, that in terms of its basic mechanism, emotional system had the same potentiality to become a system conveying referential information as the system related to actions. The classical studies of the alarm calls of vervet monkeys, as well as other studies that extended the observation to other species and other communicative contexts like social **relationship, food, inter-group aggression**, show that evolution indeed tried this pathway. This attempt was doomed, however, to failure. As noted by Hauser et al. (2002), unlike the animal examples of referential signals, most of the words of human language are not linked to a specific function (e.g. a warning cry), but "can be linked to a multiplicity of concepts". In a non-emotional communication system, the same word (e.g. fire), may indicate that fire erupted ("escape" message), but it may also indicate that the fire is ready and we can prepare the meal ("approach" message) or an almost infinite series of other meanings. In contrast, in an emotion based communicative system, a call may have a referential meaning, but, essentially, has the function of starting a specific behavior. This link with a specific response renders

the emotional system unsuitable for language evolution, in spite of its mirror mechanism.

Mirror-neuron system and sign language evolution

Monkey mirror neurons code object-directed actions. Within this limit, they solve two fundamental communication problems: parity and direct comprehension of the action. Parity requires that what count for the sender of the message count also for the receiver. Direct comprehension means that there is no need of an agreement between individuals to understand each other. No arbitrary symbols are required. The comprehension is inherent to neural organization of the individual.

The monkey mirror-neuron system is, however, a closed system linked to objects. A first problem for the mirror-neuron theory of language evolution is to explain how this close object-related system became an open system, able to describe actions and object without directly referring to them.

It is likely that the great leap from an object-related mirror-neuron system to a truly communicative mirror-neuron system is related to the development of imitation (see Arbib 2002) and the related changes observed in the human mirror-neuron system: the capacity of mirror neurons to respond to pantomimes (Buccino et al. 2001; Grèzes et al. 2003) and to discharge in response to intransitive actions (Fadiga et al. 1995; Maeda et al. 2002).

It is possible that these modifications of the mirror-neuron system did not evolve originally in order to communicate, but resulted as a consequence of the

necessity to learn, by imitation, actions made by others. Imitation implies not only the understanding of the purpose of the action to be imitated, but also the capacity to repeat the single movements that constitute an action in the right order (Tomasello and Call, 1997; Rizzolatti 2004). The necessity to keep trace of precise movements sharpened the mirror-neuron system and its capacity to convey information.

The idea that communicative actions derive from other evolutionary more ancient actions is not new. It was advanced, for example, by Van Hoof (1967) in his work on the origin of monkey communicative gestures. According to him, many of the most common communicative gestures, such as lip-smacking or lips protruded face are ritualizations of ingestive actions that monkeys use for affiliative purposes. Similarly, McNeilage (1998) suggested that the human vocal communication derived from the cyclic open-close mandibular alternation originally evolved for food ingestion. According to this view, monkey lip-smacking represents a communicative action derived from ingestive mouth movements. The existence of a neurophysiological link between ingestive and communicative actions is confirmed by the properties of F5 mouth mirror neurons (Ferrari et al. 2003, see above).

A similar notion is held for action development by Vigotski (1934), who explains pointing as derived by children attempts to grasp objects. When objects are located close to a child, the child grasps them, while, when they are located far from it, the child extends its arm and hand toward. An object-related action becomes an intransitive communicative action.

Taken together, these findings suggest that in humans (and most likely in human ancestors as well), pantomimes of actions and a variety of intransitive actions

were incorporated into the mirror-neuron system, thus acquiring its communicative properties.

We agree with Arbib (2002) that the gestural phase of communication, here just sketched, did not reach the sophisticated complexity of the modern sign language. It plausible that "protosigns" were soon accompanied by sounds and that speech development prevented the occurrence of a full-fledged sign language. The protosign language allowed, however, individuals to communicate in a much richer way than it was possible by using the emotional system. Protosigns allowed individuals to describe directions and action locations, to pantomime actions, and to give iconic descriptions of objects. Protosign language **should** have given a strong evolutionary advantage to individuals able to use it, providing in this way a strong bust for further evolution of communication.

Mirror neurons and speech evolution

The protosign communication system has a great asset: its semantics is neither arbitrarily imposed nor derives from an improbable agreement among individuals. It is inherent to the gestures that are used to communicate. This is, of course, not so for speech, or, at least, it is not so apparently. Indeed, one of the most difficult aspects of speech evolution is to provide a satisfactory answer to the problem of how words started to signify things.

Historically, the discussion on how this occurred is centered on the possible relations between the sound of a word and its meaning. On one side, there are those

who postulate a remote, but "natural" origin of the words, on the other those who regard the faculty of speech "with almost superstitious veneration, and, emulating the etymologists of Socrates, are content to ascribe the first words to arbitrary choice of gods" (Critchley, 1939, p. 15). Even if "gods" are substituted with a more prosaic concept as general agreement among speakers, the hypothesis of a natural origin of language seems to be the only one intellectually satisfactory.

A major problem with the "natural" theory is, however, the difficulty to specify what is the evolutionary link between the sound of a word and its meaning. Onomatopeia, that is the similarity between sound of the words and the sound produced by natural events or actions, done by humans or animals, is one of the suggested possibilities. Another possibility is represented by interjectional utterances emitted by individuals in certain conditions. The problem with both these hypotheses is that they are able to explain a very limited number of words. Thus, although they trace the origin of some words, they lack the generality necessary to explain the link between sound and word for most of them.

An interesting alternative hypothesis was advanced by Paget (1930). According to him, the original human communication was gestural. However, as the individual gesticulated with his hand, "his tongue, lips and jaw unconsciously followed suit in a ridiculous fashion, understudying the action of the hands". Later, the gesticulating individuals discovered that the expiration of the air through the oral cavities produced audible gestures, that is, voiced speech.

Paget gives many examples of parallelism between sound and meaning in a variety of languages. For example, as far as vowels are concerned, he suggests that

"A" (as in large) refers to anything that is large, wide open, spacious; "I" (as in mini) to something that is small or pointed; "AW" connotes a cavity (e.g yawn) and "OO" something tubular or elongated. Consonants also convey gestural symbolisms. "M" implies a continued closure; "R" implies a bending back; "DR", "TR" denotes running or walking, the direction of the tongue movement being inwards towards the speaker.

According to this theory, called "schematopoeia", the great majority of words appear to be pantomimic. They are built "much as the Chinese ideographs are, by addition of separately significant elements" (Paget 1930). This type of organization explains why it is difficult to discover the original sound meaning in the words. Furthermore, almost every action or idea can be pantomimed in many different ways and every gesture can be construed in many different ways. This account for another aspect of speech that was considered evidence against a "natural" origin of language: the dissimilarity among languages. Yet, in spite of this, similar word can be found in unrelated languages such archaic Chinese and Sumerian and other words that are dissimilar can be deciphered using the basic semantic values given to vowels and consonants.

It is obvious that the schematopoeia theory is essentially a speculation. Yet, its basic notion that the hand/ body gestures and the primitive speech gestures were intrinsically linked is very interesting. On the one side, it provides a possible clue on how messages intrinsically known (hand gestures) were transferred to an opaque gestural system, as the oro-laryngeal system, without loosing their intrinsic (non-arbitrary) meaning; on the other, a clear neurophysiological prediction derives from

it: hand/arm and speech gestures must be strictly linked and have, at least partially, a common neural substrate.

A series of recent studies demonstrates that this prediction is true. TMS experiments showed that the right *hand* motor excitability increases during reading and spontaneous speech (Tokimura et al. 1996; Seyal et al. 1999; Meister et al. 2003). The effect is limited to the left hemisphere. No language-related effects are found in the motor area of the leg. Meister et al. (2003) stressed that the increase of hand motor cortex excitability cannot be due to word articulation, because while word articulation recruits motor cortex bilaterally, the observed activation was strictly limited to the left hemisphere. The facilitation appears, therefore, to result from a co-activation of the right hand motor area with the language network.

Gentilucci and coworkers (2001) reached similar conclusions using a completely different approach. In a series of behavioral experiments, they asked participants to grasp two objects of different size with their mouths and, simultaneously, to open their right hands. The results showed that the maximal finger aperture and time to maximal finger aperture increased when the mouth was directed to the large object.

Even more relevant to the view of a strict link between hand actions and oro-laryngeal gestures is another experiment of the same study (Gentilucci et al. 2001). Participants were presented with two 3-D objects, one large, and the other small. On the visible face of the objects either two symbols or a series of dots randomly scattered on the same area occupied by the symbols were written. Participants were required to grasp the objects, but, in the condition in which the symbols appeared on

the object, they had to open their mouth. The kinematics of hand, arm, and mouth movements was recorded. The results showed that, although participants were instructed to keep the mouth aperture constant in all conditions, lip aperture and the peak velocity of lip aperture increased when the movement was directed to the large object. Control experiments showed that the effect was specific to movements of the mouth and of the contralateral hand movements. Simultaneous extension of the contralateral forearm was not affected by the main task.

In a further experiment the same authors adopted the same experimental procedure described above, but asked the participants to pronounce a syllable (e.g. GU, GA) instead of simply opening their mouth. The syllables were written on the object in the same location where the symbols appeared in the previous experiment. It was found that lip aperture was larger when the participants grasped a larger object. Furthermore, the maximal **voice** power recorded during syllable emission was also higher when grasping the larger object (Gentilucci et al. 2001).

It is clear from these experiments that both simple buccal movements and the oro-laryngeal synergies necessary for syllable emission are linked to manual gestures. Most importantly, hand actions requiring large movements share neural organization with oro-laryngeal movements coding large mouth movements. This is reminiscent of the claim that vowel A describes something large, while the vowel I (EE) something small.

Grasping movements influence syllable pronunciation not only when executed but also when observed. Participants were required to pronounce the syllables BA or GA while observing another individual grasping objects of different

size. It was found that the kinematics of lip aperture and the amplitude spectrum of voice were influenced by the observation of grasping movements done by another individual. Specifically, both lip aperture and voice peak amplitude were greater when the observed action was directed to larger objects. Control experiments ruled out that the effect was due to the size of the object or to the velocity of the observed arm movement (Gentilucci 2003).

Finally, evidence for a link between gesturing and speech system also comes from clinical studies. Hanlon et al. (1990) showed that, in aphasics, pointing with the right hand to a screen where objects are presented facilitates object naming. Similarly, Hadar et al. (1998) found that word retrieval is facilitated through gesturing in brain damaged patients.

It is obvious that the reviewed experiment by no means prove the schematopoeia theory. Yet, they indicate (those of Gentilucci in particular) that the theory is not so bizarre as one may think and that the link between hand gestures and speech system is extremely strong also in the extant *homo sapiens.*

Auditory mirror systems

Let us accept that the meaning of manual gestures, "naturally" understood through the mirror neuron mechanism, transferred, at a certain point of evolution, to oro-laryngeal gestures and that this transfer marked the beginning of language based on sound. Is this assumption plausible?

The presence of audio-visual mirror neurons in old-world monkeys (Kohler et al. 2002, see also above) suggests that auditory access to action representation is a feature common to many primate species. It is rather likely, therefore, that a link between auditory stimuli and action representation was also present in the primates from whom *Homo sapiens* descended. Thus, before speech occurrence, the pre-condition for the transfer of gesture meaning from visual to auditory modality was already present

The monkey audio-visual mirror neurons code, however, only object-related actions. Their function is similar, with the addition of auditory responses, to "classical" mirror neurons. These neurons, however, as discussed above, are not sufficient to create an efficient intentional communicative system. To achieve this, meaningful sounds (and later in evolution words) should stem, therefore, also from the association of sounds with intransitive actions and pantomimes done with the mouth and with the arms.

We already discussed some possibilities on how this visuo-auditory transfer may have occurred. Let us examine now its consequences for the cortical organization. An example may serve to this purpose.

When we eat, we move our mouth, tongue, and lips in a specific manner. The observation of this combined series of motor actions constitutes a gesture whose meaning is transparent: "eat". If, while making this action, we blow air through the oro-laryngeal cavities, we produce a sound like "mnyam-mnyam", or "mnya-mnya", words whose meaning is almost universally recognized (Paget, 1930). Such a

mechanism allows the transfer of the meaning of an action, "naturally" understood, to a sound.

The understanding of words related to oro-laryngeal gestures (like "mnyam-mnyam"), should have initially occurred through activation of audio-visual mirror neurons related to ingestive behavior or to mouth movements accompanying hand gestures. The fundamental step toward speech acquisition was achieved, however, only when individuals, thanks to their improved imitation capacities, became free to generate the sounds of actions without actually performing those actions. In analogy to what occurred to the classical mirror neuron sytem, this evolutionary step should have been accompanied by the acquisition, by the motor neurons controlling oro-pharyngeal gestures, of the capacity to resonate to sound emitted by a similar motor gestures. Thus, a new type of mirror neuron controlling sound emission and responding to meaningful sound should have been generated ("echo mirror neurons"). The incredibly confused organization of human Broca's area, where phonology, semantics, hand actions, ingestive actions, and syntax as well, are all intermixed in a rather restricted neural space (Bookheimer, 2002) is probably a consequence of this evolutive trend

Is there any evidence that humans possess an echo mirror-neuron system, i.e., a system that motorically "resonates" when the individual listen to verbal material? Recent evidence shows that is the case.

Fadiga et al. (2002) recorded MEPs from the tongue muscles in normal volunteers instructed to listen carefully to acoustically presented verbal and non-verbal stimuli. The stimuli were words, regular pseudo-words, and bitonal sounds.

In the middle of words and pseudo-words either a double "f" or a double "r" were embedded. "F" is a labio-dental fricative consonant that, when pronounced, requires slight tongue movements, while "r" is linguo-palatal fricative that, in contrast, requires, a marked tongue muscles involvement to be pronounced. During the stimulus presentation the participants' left motor cortex was stimulated with single pulse TMS.

The results showed that listening to words and pseudo-words containing the double "r" determines a significant increase of MEPs' amplitude recorded from tongue muscles with respect to listening to bitonal sounds, and words and pseudo-words containing the double "f". Furthermore, the facilitation due to the listening of the "r" consonant was stronger for word than for pseudo-words (Figure 2).

Results congruent with those of Fadiga were obtained by Watkins et al. (2003). By using single pulse TMS technique, they recorded MEPs from a lip (*orbicularis oris*) and a hand muscle (first *interosseus*), respectively, in four conditions: listening to continuous prose, listening to non-verbal sounds, viewing speech-related lip movements, and viewing eye and brow movements. Compared to control conditions, listening to and viewing speech enhanced the MEPs' amplitued recorded from the *orbicularis oris* muscle. This increase was seen only in response to stimulation of the left hemisphere. No changes of MEPs in any condition were observed following stimulation of the right hemisphere. Finally, the size of MEPs elicited in the first *interosseus* muscle did not differ in any condition.

Taken together, these data undoubtedly show that an echo mirror-neuron system exists in humans: when an individual listen to verbal stimuli there is an

automatic activation of the speech-related motor centers. More difficult is to specify the precise functional meaning of this system and its role in word understanding.

A first possibility is that the echo mirror-neuron system evolved exclusively for production/imitation of verbal sounds and still is used only for this purpose. Another hypothesis, not in contrast with the first one, is that the echo mirror-neuron system subserves, besides imitation, speech "perception" ("motor theory of speech perception" Liberman et al. 1967; Liberman and Mattingly 1985; Liberman and Wahlen 2000). According to this theory, the ultimate constituents of speech are not sounds, but articulatory gestures that evolved for the service of language. From this perspective, the echo-mirror neuron system would represent the neural mechanism transforming verbal sounds into the motor representation of the corresponding articulatory gestures.

It is important to note that the understanding of word semantics is not part of Liberman's theory. On the contrary, the evolutionary scenario, sketched above, predicts also a semantic role for the echo mirror-neurons. According to it, there are two roots to semantics: one, more ancient, based on the auditory (verbal material) activation of mirror neurons coding actions (see Pulvermueller, 2002), the other, evolutionary more recent, based on the activation of echo mirror-neurons.

This second root implies a rather interesting concept: a *second order action representation*. Classical mirror neurons do not require, in order to be triggered, action execution. The observed action is understood without its actual execution. The mirror neurons activation *represents* by itself the action (*first order action*

representation). We propose that a similar process takes place during the activation of echo mirror neurons.

The hypothesis is the following. The echo mirror neurons become active in response to verbal material and their activation evokes the motor representation of the corresponding articulatory gestures. Thus, in the case of "mnyam-mnyam", this sound recruits the corresponding articulatory representation. There is, however, something more. Because the echo-mirror neurons (e.g., those activated by "mnyam-mnyam"), because of their evolution origin must be connected with the classical mirror neurons coding the corresponding actions (e.g. to eat), when they discharge the perceiving individual recognizes not only the correct sound of the word but also its meaning.

In more general terms, at least in a competent speaker, the activity of the echo mirror neurons is sufficient by itself to represent the content of mirror neurons linked to them, without the absolute necessity of mirror-neuron system activation (and even less so of motor neurons controlling action execution). The echo-mirror neuron activation *represents*, therefore, albeit indirectly, an action (*second order action representation*). This second order representation, located in a circumscribed neural space, has clear advantage on the first order representation. It has the power to create, for example, new associations between word based on the probability of word occurrence rather than on the occurrence of actions and gives, in this way, much higher communicative possibilities to individual possessing it than those achievable with the ancient semantic system based on evocation of motor representations.

It is obvious that simple words, like the one discussed in the previous example, are rather rare, and a more complex system associating phonology to semantics **should** have evolved to link the sound of polysyllabic words (and the grammatical variations of all words) to meaning. We think, however, that the basic idea, here presented, that the echo-mirror neuron system represents a link between phonology and semantics has a strong heuristic value, and, being based on neurophysiological findings, can undergo experimental testing, and, possibly, bridge linguistic data with neurophysiology.

References

Arbib MA 2002 Beyond the mirror system: imitation and evolution of langauge. *In imitation in animals and artifacts*, ed. C Nehaniv and K Dautenhan, pp. 229-280. Cambridge MA: The MIT Press.

Armstrong AC, Stokoe WC, Wilcox SE. 1995. *Gesture and the nature of language.* Cambridge, UK: Cambridge University Press.

Bookheimer, S. 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Ann. Rev. Neurosci. 25:* 151-88

Brodmann, K (1909). Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien Dargerstellt auf Grund des Zellenbaues. Leipzig: Barth.

Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L et al. 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13:400-4.

Buccino G, Vogt S, Ritzl A, Fink GR, Zilles K et al. 2004. Neural circuits underlying imitation of hand actions: an event related fMRI study. *Neuron,* in press

Byrne RW. 1995. *The Thinking Ape. Evolutionary Origins of Intelligence.* Oxford, U.K.: Oxford University Press.

Carr L, Iacoboni M, Dubeau MC, Mazziotta JC, Lenzi GL. 2003. Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl Acad. Sci. U.S.A. 100,* 5497-5502.

Corballis MC. 2002. *From hand to mouth. The origins of language.* Princeton: Princeton University Press.

Critchley M. 1939. *The language of gesture.* London: Edward Arnold & Co..

Fadiga L, Fogassi L, Pavesi G, Rizzolatti G. 1995. Motor facilitation during action observation: A magnetic stimulation study. *J. Neurophysiol.* 73:2608-11.

Fadiga L, Craighero L, Buccino G, Rizzolatti G. 2002. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15:399-402

Ferrari PF, Gallese V, Rizzolatti G, Fogassi L. 2003. Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *Eur. J. Neurosci.* 17:1703-14.

Fogassi L, Gallese V, Fadiga L, Rizzolatti G. 1998. Neurons responding to the sight of goal directed hand/arm actions in the parietal area PF (7b) of the macaque monkey. *Soc. Neurosci.* 24:257.5 (Abstr.)

Galef BG. 1988. Imitation in animals: history, definition and interpretation of data from psychological laboratory. In *Comparative social learning,* ed. T Zental, BG Galef, pp. 3-28, Hillsdale, NJ: Erlbaum.

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. 1996. Action recognition in the premotor cortex. *Brain* 119:593-609.

Gallese V, Fogassi L, Fadiga L, Rizzolatti G. 2002. Action representation and the inferior parietal lobule. In *Attention & Performance XIX. Common mechanisms in perception and action,* ed. W Prinz, B Hommel. Oxford: Oxford University Press.

Gangitano M, Mottaghy FM, Pascual-Leone A. 2001. Phase specific modulation of cortical motor output during movement observation. *NeuroReport* 12:1489-92.

Gentilucci M, Benuzzi F, Gangitano M, Grimaldi S. 2001. Grasp with hand and mouth: a kinematic study on healthy subjects. *J. Neurophysiol.* 86:1685-99.

Gentilucci M 2003 Grasp observation influences speech production. *Eur. J. Neurosci.* 17:179-84.

Grèzes J, Armony JL, Rowe J, Passingham RE. 2003. Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage* 18:928-37.

Hadar U, Wenkert-Olenik D, Krauss R, Soroker N. 1998. Gesture and the processing of speech: neuropsychological evidence. *Brain and Language* 62:107-26.

Hanlon RE, Brown JW, Gerstman LJ. 1990. Enhancement of naming in nonfluent aphasia through gesture. *Brain and Language* 38:298-314.

Hauser MD, Chomsky N, Fitch WT. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science* 22; 298:1569-79.

Heiser M, Iacoboni M, Maeda F, Marcus J, Mazziotta JC. 2003. The essential role of Broca's area in imitation. *Eur. J. Neurosci.* 17:1123-8.

Hyvarinen J. 1982. Posterior parietal lobe of the primate brain. *Physiol. Rev.* 62: 1060-129.

Iacoboni M, Woods RP, Brass M, Bekkering H, Mazziotta JC, Rizzolatti G. 1999. Cortical mechanisms of human imitation. *Science* 286:2526-8.

Iacoboni M, Koski LM, Brass M, Bekkering H, Woods RP et al. 2001. Reafferent copies of imitated actions in the right superior temporal cortex. *Proc. Natl. Acad. Sci.* 98:13995-9.

Jeannerod M. 1994. The representing brain. Neural correlates of motor intention and imagery. *Behav. and Brain Sci.* 17:187-245.

Jellema T, Baker CI, Wicker B, Perrett DI. 2000. Neural representation for the perception of the intentionality of actions. *Brain Cogn.* 442:280-302.

Jellema T, Baker CI, Oram MW, Perrett DI. 2002. Cell populations in the banks of the superior temporal sulcus of the macaque monkey and imitation. In *The imitative mind. Development, evolution and brain bases,* ed., AN Melzoff, W Prinz. Cambridge: Cambridge University Press.

Jürgens, U. 2002. Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.*, **26**, 235-258.

Kohler E, Keysers C, Umiltà MA, Fogassi L, Gallese V, Rizzolatti G. 2002. Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science* 297:846-48.

Koski L, Wohlschlager A, Bekkering H, Woods RP, Dubeau MC. 2002. Modulation of motor and premotor activity during imitation of target-directed actions. *Cereb. Cortex* 12:847-55.

Krolak-Salmon P, Henaff MA, Isnard J, Tallon-Baudry C, Guenot M, Vighetto A, Bertrand O, Mauguiere F (2003) An attention modulated response to disgust in human ventral anterior insula. *Ann. Neurol.* 53: 446-453.

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. 1967 Perception of the speech code. *Psychol. Rev.* 74:431-61.

Liberman AM, Mattingly IG. 1985 The motor theory of speech perception revised. *Cognition* 21:1-36.

Liberman AM, Whalen DH. 2000 On the relation of speech to language. *Trends Cogn. Neurosci.* 4:187-96.

MacNeilage PF. 1998. The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 21:499-511.

Maeda F, Kleiner-Fisman G, Pascual-Leone A. 2002. Motor facilitation while observing hand actions: specificity of the effect and role of observer's orientation. *J. Neurophysiol.* 87:1329-35.

Meister IG, Boroojerdi B, Foltys H, Sparing R, Huber W, Topper R. 2003. Motor cortex hand area and speech: implications for the development of language. *Neuropsychologia* 41:401-6.

Nishitani N, Hari R. 2000. Temporal dynamics of cortical representation for action. *Proc Natl Acad Sci U S A* 97:913-8.

Nishitani N, Hari R. 2002. Viewing lip forms: cortical dynamics. *Neuron* 36:1211-20.

Paget R. 1930. *Human Speech.* London: Kegan Paul, Trench and Co..

Perrett DI, Harries MH, Bevan R, Thomas S, Benson PJ et al. 1989. Frameworks of analysis for the neural representation of animate objects and actions. *J. Exp. Biol.* 146:87-113.

Perrett DI, Mistlin AJ, Harries MH, Chitty AJ. 1990. Understanding the visual appearance and consequence of hand actions. In *Vision and Action: The Control of Grasping,* ed. MA Goodale, pp. 163-342. Norwood, NJ: Ablex.

Pulvermueller F. 2002. *The neuroscience of language.* Cambridge, UK: Cambridge University Press. pp 315.

Rizzolatti G, Fadiga L, Fogassi L, Gallese V. 1996. Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3:131-41.

Rizzolatti G, Arbib MA. 1998. Language within our grasp. *Trends Neurosci.* 21:188-94.

Rizzolatti G, Fogassi L, Gallese V. 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2:661-70.

Rizzolatti G 2004. The mirror-neuron system and imitation. In *Perspectives on Imitation: From Mirror Neurons to Memes*, ed S Hurley and N Chater. Cambridge MA: MIT Press.

Seyal M, Mull B, Bhullar N, Ahmad T, Gage B. 1999. Anticipation and execution of a simple reading task enhance corticospinal excitability. *Clin Neurophysiol.* 110:424-9.

Tanaka, S., and Inui, T. (2002). Cortical involvement for action imitation of hand/arm postures versus finger configurations: an fMRI study. Neuroreport *13,* 1599-1602.

Tokimura H, Tokimura Y, Oliviero A, Asakura T, Rothwell JC. 1996. Speech-induced changes in corticospinal excitability. *Ann Neurol.* 40:628-34.

Tomasello M, Call J. 1997. *Primate Cognition* Oxford: Oxford University Press.

Van Hoof JARAM 1967. The facial displays of the catarrhine monkeys and apes. In *Primate Ethology*, ed. D Morris pp.7-68 London: Weidenfield & Nicolson.

Vygotsky LS. 1934. *Thought and language.* Cambridge, MA: MIT Press.

Visalberghi E, Fragaszy D. 2001. Do monkeys ape? Ten years after. In *Imitation in animals and Artifacts,* ed. K Dautenhahn, C Nehaniv. Boston, Ma: MIT Press.

Von Bonin G, Bailey P. 1947. *The neocortex of macaca mulatta.* Urbana, IL: University of Illinois Press, 136 pp.

Von Economo C. 1929. *The Cytoarchitectonics of the Human Cerebral Cortex.* London: Oxford University Press. 186 pp.

Watkins KE, Strafella AP, Paus T. 2003. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41:989-94.

Whiten A. and Ham R. 1992. On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. In *Advances in the study of behavior,* ed. PBJ Slater, JS Rosenblatt, C Beer and M Milinski, pp 239-83. San Diego: Academic Press.

Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G. 2004. Both of us disgusted in *my* insula: the common neural basis of seeing and feeling disgust *Neuron*, in press.

Figure legends

Figure 1 Examples of two communicative mirror neurons. Left side (neuron 76). Activity during: A, observation of lip smacking; B, observation of lips protruding; C, observation of sucking a syringe. All gestures were made by an experimenter. D, discharge associated with monkey's active food ingestion accompanied by lips protruding. Right side (neuron 28). Activity during: A, observation of lips protruding; B, observation of holding food with the mouth; C, food presentation. The gestures were made by an experimenter. D, discharge associate with monkey's active sucking juice from a syringe. Ordinates : spikes/s; Abscissae: time, bin width 20 ms. (From Ferrari et al. 2003).

Figure 2. Normalized motor evoked potential (MEPs) areas recorded from tongue muscles during listening to words, psudo-words and bitonal sounds. Data from all subjects. "rr" refers to verbal stimuli containing lingua-palatal fricative consonants; "ff" refers to verbal material containing labio-dental fricative consonants.