Language within our grasp

Giacomo Rizzolatti and Michael A. Arbib

In monkeys, the rostral part of ventral premotor cortex (area F5) contains neurons that discharge, both when the monkey grasps or manipulates objects and when it observes the experimenter making similar actions. These neurons (mirror neurons) appear to represent a system that matches observed events to similar, internally generated actions, and in this way forms a link between the observer and the actor. Transcranial magnetic stimulation and positron emission tomography (PET) experiments suggest that a mirror system for gesture recognition also exists in humans and includes Broca’s area. We propose here that such an observation/execution matching system provides a necessary bridge from ‘doing’ to ‘communicating’, as the link between actor and observer becomes a link between the sender and the receiver of each message.


‘In all communication, sender and receiver must be bound by a common understanding about what counts; what counts for the sender must count for the receiver, else communication does not occur. Moreover the processes of production and perception must somehow be linked; their representation must, at some point, be the same.’

What is said here by Alvin Liberman for speech where individuals have an explicit intent to communicate, may apply also for ‘communications’ in which such an overt intention is absent. We understand when one individual is attacking another or when someone is peacefully eating an apple. How do we do it? What is shared by the (involuntary) sender and by the receiver? Is this mechanism the precursor of willed communications? The present review addresses these questions.

The mirror system

Neurons located in the rostral part of monkey inferior area 6 (area F5) discharge during active movements of the hand or mouth, or both8,9. Some years ago we found that in most F5 neurons, the discharge correlates with an action, rather than with the individual movements that form it10. Accordingly, we classified F5 neurons into various categories corresponding to the action associated with their discharge. The most common are: ‘grasping with the hand’ neurons, ‘holding’ neurons and ‘tearing’ neurons6,11. Further study revealed something unexpected: a class of F5 neurons that discharge not only when the monkey grasped or manipulated the objects, but also when the monkey observed the experimenter making a similar gesture8. We called the neurons endowed with this property ‘mirror neurons’ (Fig. 1).

The response properties of mirror neurons to visual stimuli can be summarized as follows: mirror neurons do not discharge in response to object presentation; in order to be triggered they require a specific observed action. The majority of them respond selectively when the monkey observes one type of action (such as grasping). Some are highly specific, coding not only the action aim, but also how that action is executed. They fire, for example, during observation of grasping movements, but only when the object is grasped with the index finger and the thumb.

All mirror neurons show visual generalization: they discharge when the agent of the observed action (typically a hand) is far away from or close to the monkey. A few neurons respond even when the object is grasped by the mouth. The actions most represented are: grasp, manipulate, tear, and put an object on a plate. Mirror neurons also have motor properties that are indistinguishable from those of F5 neurons that do not respond to action observation. In this review, they will be referred to collectively and regardless of their other properties, as ‘canonical neurons’. Typically, mirror neurons show congruence between the observed and executed action. This congruence can be extremely strict, that is, the effective motor action (for example, precision grip) corresponds with the action that, when seen, triggers the neuron (that is, precision grip). For other neurons the congruence is broader: the motor requirements (for example, precision grip) are usually stricter than the visual ones (for example, any type of hand grasping). An example of a highly congruent mirror neuron is shown in Fig. 2. What is
the function of mirror neurons? The proposal that we7,8 and others9 have advanced is that their activity ‘represents’ actions. This representation can be used for imitating actions and for understanding them. By ‘understanding’ we mean the capacity that individuals have to recognize that another individual is performing an action, to differentiate the observed action from other actions, and to use this information to act appropriately. According to this view, mirror neurons represent the link between sender and receiver that Liberman postulated in his motor theory of speech perception as the necessary prerequisite for any type of communication1,10,11.

What is area F5?

Although doubts have been expressed12, most authors share the view that the rostral part of the monkey ventral premotor cortex (area F5) is the monkey homolog of Broca’s area in the human brain. The reasons for this view are: that both F5 and Broca’s area are parts of inferior area 6 (Refs 13–15) and their location within the agranular frontal cortex is similar (Box 1); and cytoarchitectonically, there are strong similarities between area 44 (the caudal part of Broca’s area) and F5 (Refs 14,16,17).

Functionally, a difference between Broca’s area and F5 is that Broca’s area is most commonly thought of as an area for speech, whereas F5 is often considered as an area for hand movements. F5 is somatotopically organized – its dorsal part contains a representation of hand movements14,15, and its large ventral part contains a representation of mouth and larynx movement14,15; a similar organization is present in the ventral premotor cortex of other primates21. Similarly, the motor properties of human Broca’s area do not relate only to speech: recent PET data indicate that Broca’s area might also become active during the execution of hand or arm movements34,35, during mental imagery of hand grasping movement (mostly area 44)36,37, and during tasks involving hand–mental rotations (areas 44 and 45)38. Finally, Broca’s area becomes active in patients who have recovered from subcortical infarctions when they are asked to use their paralyzed hand39.

It is intriguing that the area, which in the monkey contains a system that links action recognition and action production, is precisely that area that, for completely different reasons, has been proposed as the homologue of Broca’s area. Is this a mere coincidence?

Or, on the contrary, has the mirror system been fundamental for the development of speech and, before speech, of other forms of intentional communication? Before discussing these points, we examine the evidence for a mirror system in humans.
The mirror system in humans

The first demonstration of a mirror system in humans was provided by Fadiga et al.18 The rationale of their experiment was the following: if the observation of an action activates the premotor cortex in humans as it does in monkeys, then magnetic transcranial stimulation should induce, during action observation, an enhancement of motor-evoked potentials recorded from those muscles that are active when the observed action is executed. Their results confirmed the hypothesis: during the observation of various actions, a selective increase of motor evoked potentials occurred in the muscles that the subjects usually use for producing them.

Although these data indicate that an action production and action observation matching system exists in humans, they do not give information on the circuits that underlie it. Data on this issue were provided by two PET experiments29,30. The two experiments differed in many aspects, but both had a condition in which subjects observed the experimenter grasping a 3-dimensional object; object observation was used as a control situation. Grasp observation significantly activated the superior temporal sulcus (STS), the inferior parietal lobule and the inferior frontal gyrus (area 45); all activations were in the left hemisphere.

Hemispheric differences aside, the cortical areas active during action observation in humans match well with those active in the monkey under the same conditions. Neurons that become selectively active in the STS during the sight of hand actions were described by Perrett and his co-workers28,31. There is evidence, although limited, that mirror neurons might be present in area 7b in monkeys (Ref. 32). Finally, as discussed earlier, it is likely that F5 is the monkey homolog of Broca’s area.

Taken together, human and monkey data indicate that, in primates, there is a fundamental mechanism for action recognition. We argue that individuals recognize actions made by others because the neural pattern elicited in their premotor areas during action observation is similar to that internally generated to produce that action. This mechanism in humans is circumscribed to the left hemisphere. In the next section we will posit that this action-recognition mechanism has been the basis for language development.

Action recognition and communication

Animals’ calls and human speech are undoubtedly different phenomena. Among the many aspects that differentiate them is a marked difference in the anatomical structures underlying the two behaviors. Animals calls are mediated primarily by the cingulate cortex plus some diencephalic and brain stem structures12,13. Speech is mediated essentially by a circuit whose main nodes are the classical Broca’s and Wernicke’s areas, both located on the lateral cortical surface.

Our proposal is that the development of the human lateral speech circuit is a consequence of the fact that the precursor of Broca’s area was endowed, before speech appearance, with a mechanism for recognizing actions made by others. This mechanism was the neural prerequisite for the development of interindividual communication and finally of speech. We thus view language in a more general setting than one that sees speech as its complete basis.
system will allow a brief prefix of the movement to be exhibited. This prefix will be recognized by the other individual. This fact will affect both the actor and the observer. The actor will recognize an intention in the other individual. This dialogue forms the core of language. The capacity to notice that one has emitted a signal and associating this with changes of the behavior of others might or might not have developed simultaneously. However, there is no doubt that, once established, this new association should have yielded enormous benefits of adaptive value for the group of individuals that started to make use of it, providing the selective pressure for the extension of communicative capacities to larger groups. This new use of the mirror system, at both individual and species levels, marks the beginning of intentional communication. What actions were used for this new function in primates? Hand gestures or oro-facial movements? Before examining this issue, it is necessary to examine whether or not a "prelinguistic grammar" can be assigned to the control and observability of actions. If this is so, the notion that evolution could yield a language system ‘atop’ of the action system becomes much more plausible.

---

**Box 1. Cytoarchitectonic map of the caudal part of the monkey frontal lobe and possible homologies with human frontal cortex**

Figure A shows parcellation of prearcuate cortex and agranular frontal cortex of the macaque monkey and Fig. B shows parcellation of the region of the human frontal cortex defined as ‘intermediate precentral cortex’ by Campbell. The terminology of Foerster and Vogt has been adopted for the human cortex. Similar colors in A and B indicate areas with anatomical and functional homologies. Brain regions colored yellow are areas with anatomical and functional homologies, mostly related to orienting behavior; areas colored red also share anatomical and functional homologies and are mostly related to interactions with the external world.4-6

The homology is based on cytoarchitectonics, electrical stimulation, and sulci embryology.

<table>
<thead>
<tr>
<th>References</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Walker, A.E. (1940)</td>
<td>J. Comp. Neurol. 82, 256–270</td>
</tr>
<tr>
<td>Foerster, O. (1936)</td>
<td>Brain 59, 135–159</td>
</tr>
</tbody>
</table>
A pre-linguistic 'grammar' of action in the monkey brain

In order to provide abstract expression of the 'meanings' of neural activity in premotor cortex (monkey area F5), we have chosen 'case grammar' as a representation of sentence structure. Case grammar organizes sentences around action frames with slots for different roles. The key paper for case structure is 'The Case for Case' by Fillmore36, although many of the ideas about case have now been absorbed in the thematic structure of the lexicon that is an integral component of the Chomskian approach to syntax known as 'government and binding theory'37. In a case analysis, the sentence 'John hit Mary with his hand' is viewed as the 'surface structure' for a case structure 'hit' (John, Mary, John's hand), which is an instance of the case frame 'hit' (agent, recipient, instrument), which makes explicit the roles of 'John', 'Mary' and 'John's hand'. Clearly, many different sentences can express the underlying case structure. Our analysis will deal with the two main types of F5 neurons: the mirror neurons and the 'canonical' F5 neurons.

(1) Imperative structure for 'canonical' F5 neurons. We view the activity of 'canonical' F5 neurons as part of the code for an imperative case structure, for example,

Command: grasp-A(raisin)
as an instance of grasp-A(object), where grasp-A is a specific kind of grasp, to be applied to the raisin. Note that this case structure is an 'action description', not a linguistic representation. 'raisin' denotes the specific object towards which the grasp is directed, whereas grasp-A is a specific command directed towards an object with well-specified physical properties, but whose semantic properties are unspecified. The idea is that grasp type generalizes across similar grasps of varied objects has been postulated in opposition space theory38. Note that the slots in a case frame come with restrictions on what can fill those slots, for example, any x in grasp-A (x) must be a small object. As actions become more refined and as the transition to language occurs, the constraints on slot fillers might become more rigorous. From this, it follows that if the same principle holds for linguistic commands as for motor commands, Broca's area would code 'verb phrases' and constraints on the noun phrases that can fill the slots, but not details of the noun phrases themselves. This knowledge (objects or noun phrases) could be completely outside F5 or Broca's area; for example, in the temporal lobe.

(2) Declarative structure for mirror neurons. Having viewed the activity of canonical F5 neurons as coding a command (compared with an imperative sentence), we might say that the firing of 'mirror' F5 neurons is part of the code for a declarative case structure, for example,

Declaration: grasp-A Luigi, raisin)

which is a special case of grasp-A(agent, object), where grasp-A is a specific kind of grasp, applied to the raisin (the object) by Luigi (the agent). Again, this is an 'action description', not a linguistic representation. If attention is focused on the agent's hand, then the appropriate case structure would be grasp-A (hand, object) as a special case of grasp-A (instrument, object). Thus, the same act can be perceived in different ways: 'whos' grasps versus 'what' grasps. An interesting aspect of mirror-neuron properties is that they do not fire when the monkey observes the experimenter grasping the raisin with pliers, rather than his hand (Fig. 1B). However, after repetitive observation, a response to the tool may appear (G. Rizzolatti, unpublished).

We thus see the ability to learn new constraints on a case slot: in this case the observed generalization of the 'instrument' role from hands alone to include pliers.

(3) Declarative structure for non-mirror 'canonical' F5 neurons. In a case of grasp-A (object), once the grasp is initiated, can it be asserted that the activity in F5 now becomes part of the declarative 'grasp-A (self, object)'? This is an even more refined command. For these neurons, the discharge that accompanies the hand movement might have a specific declarative function concerning the agent of the action. The sentence in this case would be 'grasp-A (self, object)'.

From action to speech

Returning to our previous query as to which was the initial communicative gestural system in primates, we now distinguish between systems that are closed in the sense that they have a small, fixed repertoire and systems whose elements can be combined to yield an open repertoire of meaning. We argue that: (1) the mimetic capacity inherent to F5 and Broca's area had the potential to produce various types of closed systems related to the different types of motor fields present in that area (hand, mouth and larynx); (2) the first open system to evolve en route to human speech was a manual gestural system that exploited the observation and execution matching system described earlier; and (3) that this paved the way for the evolution of the open vocalization system we know as speech.

As far as the first point is concerned, both F5 and Broca's areas have the neural structures for controlling oro-laryngeal, oro-facial and brachio-manual movements. Furthermore, they are both endowed with mechanisms that link action perception and action production. This is true for brachio-manual gestures (as discussed in this review) as well as, in the case of human Broca's area, for linguistic tasks, including those not requiring speech production39–41. It is likely that the human capacity to communicate beyond that of other primates may have depended on the progressive evolution of the mirror system in its globality. Congruent with this view is the observation by Donald40 that mimetic capacity, a natural extension of action recognition, is central to human culture (such as dances, games and tribal rituals), and that the evolution of this capacity was a necessary precursor to the evolution of language. Even so, it is interesting to speculate on the sequence of events that led to the evolution of communicative speech. The gestures of primates that were most likely to be first used for person-to-person communication (as distinct from signals that are directed
to ‘everybody’, rather than to a specific receiver) are the oro-facial ones. In favor of this view are the following facts: oro-facial movements are used for communication by monkeys, apes and humans, and neither monkeys nor humans use manus-brachial gestures as their main natural way to communicate. Exceptions are deaf people who naturally use sign language, some Indian populations in North America and some Aboriginal Australian tribes. The open-close alternation of the mandible that is typical of oro-facial communication of monkeys (‘limpsacks’, ‘tonguemsacks’) appears to persist in humans where it forms the syllabic ‘frame’ in speech production.

With the addition of a sound had only an emotional valence that simply reinforced the meaning conveyed by the oro-facial communication, the exchange of communication is essentially limited to two actors. The possibility of introducing a third element in this one-to-one communication is very limited. By contrast, if manual gestures are associated with oro-facial communication, the sender’s possibilities dramatically increase. The sender might indicate to the receiver the position of a third person or of an object, or even give a certain description of their characteristics. A second reason lies in the fact that the combinatorial properties for the openness of speech are virtually absent in the basic primate form of oro-facial communication. By contrast, they are inherent to the brachio-manual system, both when it is used for transitive actions (actions directed towards objects) and when it is employed for intransitive gestures (as in the case of American Sign Language).

These considerations suggest that, at a certain stage, a brachio-manual communication system evolved complementing the oro-facial one. This development greatly modified the importance of vocalization and its control. Whereas during the closed oro-facial stage, sounds could add very little to the gestural message (for example, oro-facial gesture ‘be scared’, oro-facial gesture plus vocalization ‘be more scared’), their association with gestures allowed them to assume the more open, referential character that brachio-manual gestures had already achieved. An object or event described gesturally (such as, large object - large gesture of the arms, and small object - tiny opening of the mouth, vowel ‘a’, and small object - tiny opening of the mouth, vocal ‘i’), a primitive vocabulary of meaningful sounds could start to develop.

An important consequence of this new functional use of vocalization was the necessity of its skillful control. In the oro-facial communication system, the addition of a sound had only an emotional valence that simply reinforced the meaning conveyed by the facial expression, its precise execution had a relative importance. Therefore, vocalization could remain under the control of the old system located in the brain medial areas. The situation changed radically when sounds acquired a descriptive value and thus had to remain the same in identical situations and, in addition, had to be imitated when emitted by other individuals. These new requirements could not be fulfilled by the ancient emotional vocalization centers.

This new situation was most likely to be the cause of the emergence of human Rinca’s area from an F5-like precursor that already had mirror properties, a control of oro-laryngeal movements and, in addition, a tight link with the adjacent primary motor cortex. The evolutionary pressure for more complex (combinatorial) sound emission, and the anatomical possibility for it, were thus the elements that moved language from its manus-brachial origins to sound emission. Manual gestures progressively lost their importance, whereas, by contrast, vocalization acquired autonomy, until the relation between gestural and vocal communication inverted and gesture became purely accessory to sound communication. At this point speech took off.

A historical coda

The debate on the origin of language has a long history. Clearly we side with those authors who see a common origin for human speech and some forms of communications in primates, with gestural communication playing an important role in human language genesis.

Chomsky has long argued that language is determined by innate, biologically determined abilities in conjunction with exposure to the language in the environment; to learn a grammar the child must simply use a few fragments of a particular language to have ‘set’ parameters in the Universal Grammar, a genetically determined, biological endowment relevant to language (for a critique of this view see Ref. 58).

Chomsky also downplays the role of natural selection in language evolution. Pinker and Bloom offer convincing arguments against this view, but their approach to language evolution differs from ours on one major point. They see what has evolved for language as Universal Grammar. Our suggestion, by contrast, is that natural selection yielded a set of generic structures for matching action observation and execution. These structures, coupled with appropriate learning mechanisms, proved great enough to support cultural evolution of human languages in all their richness. We hold that human language (as well as some dyadic forms of primate communication) evolved from a basic mechanism that was not originally related to communication: the capacity to recognize actions.

Imprints in fossil cranial cavities indicate that ‘speech areas’ were already present in early hominids such as Homo habilis (Ref. 60), but there is debate over whether or not such areas were already present or not in australopithecines. A plausible hypothesis is that the transition from the australopithecines to the first forms of Homo coincided with the transition from a mirror system, enlarged, but used only for action recognition, to a human-like mirror system used for intentional communication. Our view on the subsequent scenario is close to that of Corballis. The proto-speech’ areas of early hominids mediated oro-facial and brachio-manual communication, but not speech. The long period from the appearance of these areas to the appearance of speech coincided with an increased capacity to communicate with gesture and the progressive association of gesture with vocalization.

In conclusion, the discovery of the mirror system suggests a strong link between speech and action representation. ‘One sees a distinctly linguistic way of
doing things down among the nuts and bolts of action and perception, for it is there, not in the remote recesses of cognitive machinery, that the specifically linguistic constituents make their first appearance.6

Acknowledgments


