

2 Understanding Others: Imitation, Language, and Empathy

Marco Iacoboni

2.1 Introduction

The study of the neural basis of imitation is in its first stage. Until a few years ago, the only available information on the neural underpinnings of imitative behavior was restricted to lesion data from neurological patients. Although extremely valuable, the information obtained from neurological observations is limited, mostly because the lesions causing the imitative deficits are naturally occurring ones and do not have the precise anatomical boundaries that allow a detailed study of the brain-behavior relationship. Two main factors have limited the neuroscience of imitation. First, there is little consensus on a definition of imitation (R. Byrne & Russon, 1998; Heyes, 2002). This lack of consensus has reduced the enthusiasm of neuroscientists for investigating the neural basis of imitative behavior. Second, even though some neuroimaging techniques have been around for about a quarter of a century, brain mappers initially had the tendency to stay away from complex phenomena, and imitation has been definitely perceived by them as a complex phenomenon. These two factors have led to the paradoxical situation of the late 1990s, when there were tens of peer-reviewed imaging studies on, say, saccades, and not even one on imitation!

A series of reports on experiments investigating the neural basis of imitation, however, has been published recently (see, for instance, Decety and Chaminade, vol. 1, ch. 4). This trend seems to result from two main causes. First, there is a recent tendency in the neuroimaging world to study complex phenomena, such as theory of mind (C. Frith & Frith, 1999) or even "social"-like interactions (Montague et al., 2002; Rilling et al., 2002). Second, macaque single-cell observations published in the early 1990s have provided good neuronal models of functional properties that are relevant to imitation (di Pellegrino et al., 1992; Gallese et al., 1996). This is particularly important because the mainstream imaging techniques generally rely on

indirect measures of neuronal activity, such as blood flow. The existence of neurophysiological data that can help constrain the interpretation of the imaging data is generally considered extremely valuable.

In this chapter I summarize the most meaningful data obtained so far on the neural underpinnings of imitation. The plan is to relate these findings to a neural and functional model of imitation and its relations with two other functional domains, language and empathy. The approach I use here envisions brain mapping techniques as investigative techniques with explanatory power. Typically, brain mapping is perceived as some kind of sophisticated phrenology. Detailed aspects of cognitive functions are mapped onto precise neural structures. Obviously, the map obtained looks a lot more sophisticated than the phrenological maps of the nineteenth century. However, the explanatory power of this approach remains limited with regard to testing models. What I advocate here is an approach that combines imaging data with functional information obtained from single-cell observations. With this approach, it is possible to test information-processing models of imitation and its relations with other domains.

2.2 Minimal Neural Architecture for Imitation

2.2.1 An Action Recognition System in the Macaque Brain

Two European laboratories, David Perrett's and Giacomo Rizzolatti's, have systematically studied the properties of temporal, parietal, and frontal neural systems of the macaque brain that seem relevant to action representation and potentially to imitation. Following the leads that resulted from the studies of Charles Gross on the complex visual properties of inferior temporal neurons, Perrett and his collaborators have studied neurons in the superior temporal sulcus (STS) that respond to moving biological stimuli, such as hands, faces, and bodies (Perrett et al., 1989, 1990a; Perrett & Emery, 1994). These neurons seem to respond to moving bodies and body parts only when the body or body part is engaged in goal-oriented actions. For instance, some of these neurons respond to the sight of a hand reaching and grasping an object. The same neuron will not fire at the sight of the hand reaching toward the object but not grasping it. The modulation of activity in STS neurons is independent of low-level visual features. In fact, a point-light version of the same action, that is, a hand reaching and grasping an object, is enough to activate a neuronal response in these STS cells (Jellema et al., 2002). In other words, what these STS neurons code is the sight of a meaningful interaction between an object and an intentional agent.

The properties of STS neurons are limited exclusively, at least so far, to the visual domain, in that no neuronal responses in STS seem associated with motor behavior. In contrast, Giacomo Rizzolatti and his collaborators have described frontal and parietal neurons with motor properties (in that they are active when a monkey performs a movement) that also have visual responses similar to the ones observed in STS by Perrett (di Pellegrino et al., 1992; Gallese et al., 1996). These neurons have been described for the first time in a region of the inferior frontal cortex called area F5, according to an anatomical nomenclature that is becoming increasingly used (Matelli et al., 1985). In area F5 there exist two types of neurons with identical motor properties and quite different visual properties. The two types of neurons are called *canonical* and *mirror*. Both types fire when a monkey executes goal-directed actions, such as grasping, holding, tearing, and manipulating. Some of these neurons fire for a precision grip, as when a monkey grasps small objects like a raisin, and some other neurons fire for a whole-hand grasp, as when a monkey grasps larger objects, such as an apple. When it comes to their visual properties, canonical neurons that fire when a monkey grasps a small object with a precision grip also respond to the sight of small objects that can be grasped with a precision grip, but not to the sight of larger objects graspable with, say, a whole-hand grip. Note that these visual responses also occur when a monkey does not reach and grasp the object; the simple sight of the object is sufficient to activate canonical neurons. In other words, canonical neurons seem to be coding the affordance of an object, the pragmatic aspect of how to grab it, rather than its semantic content.

In contrast, mirror neurons do not fire at the sight of an object, but will fire at the sight of a whole action. So, say that there is a neuron in F5 that fires when a monkey grasps an object. That same neuron, if it is a mirror neuron, will fire at the sight of another individual grasping an object, but will not fire at the sight of the object alone and will not fire at the sight of a pantomime of a grasp in the absence of the object. In other words, these neurons seem to be matching the execution and the observation of an action. The functional properties of these neurons suggest that they may implement a simple, noninferential mechanism of action recognition based on neural identity. This mechanism may be a building block for imitative behavior.

A posterior parietal area of the macaque, area PF, situated in the rostral sector of the inferior parietal lobule, contains mirror neurons with functional properties that are substantially identical to the ones described in F5 (Rizzolatti et al., 2001). Area PF and area F5 are anatomically connected

with robust projections (Rizzolatti et al., 1998). This pattern of cortico-cortical connectivity leads us to believe that F5 and PF belong to an integrated circuit for action recognition. Furthermore, STS, the region where Perrett has discovered the neurons with the complex visual properties described earlier, is connected with the posterior parietal cortex (Seltzer & Pandya, 1994). Thus these three cortical regions of the macaque brain, STS in the superior temporal cortex, area F5 in the inferior frontal cortex, and area PF in the posterior parietal cortex, seem to have functional properties and connectivity patterns that may instantiate a whole circuit for coding actions. The question that I address in the next section is whether there is a similar circuit for recognition of actions and possibly imitation in the human brain.

2.2.2 Minimal Human Neural Architecture for Imitation

The first attempts to demonstrate an action recognition system in the human brain similar to the one in the macaque brain were made using positron emission tomography (PET) and, as activation tasks, execution and observation of grasping (Grafton et al., 1996; Rizzolatti et al., 1996b). The idea behind these studies was the following. If there is a human action recognition system that is similar to the one described in macaques, motor areas in the human brain belonging to this system should be active during both execution and observation of grasping. Ideally, there should also be some anatomical correspondence between the human and the macaque areas. The early PET attempts were not entirely successful, even though some aspects of the empirical findings were encouraging. The two broadly defined regions of superior temporal cortex and inferior frontal cortex were indeed activated during both observation and execution of action. The areas activated within inferior frontal and superior temporal cortex during execution of grasping, however, did not spatially match the areas activated within inferior frontal and superior temporal cortex during observation of grasping. Furthermore, the posterior parietal cortex was found to be activated only during execution of grasping (Rizzolatti et al., 1996b). However, a second study comparing observation of grasping with imagination of this action did report activation of posterior parietal cortex during observation (Grafton et al., 1996). The reason the results of these first studies were successful only in part is probably due to technical limitations of the 2-D PET methodology used.

The second important feature of the action recognition system described in the macaque is that it is driven by goal-directed actions. To test whether human brain areas of the grasping circuit show a similar feature, we per-

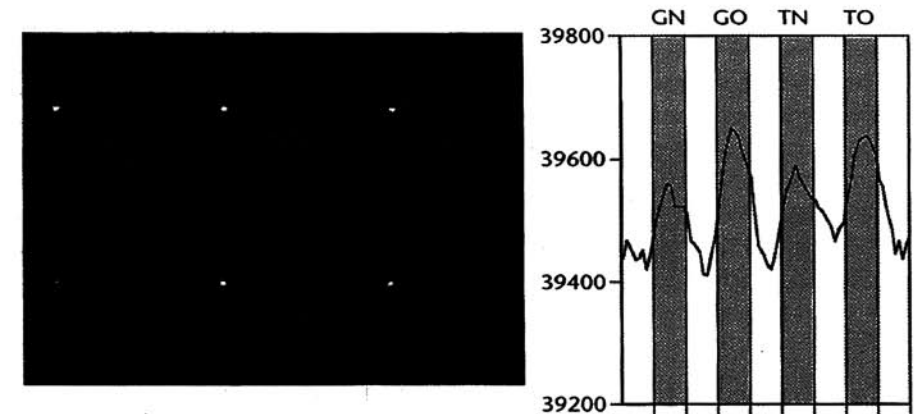


Figure 2.1

Activity in the human inferior frontal cortex (putative Brodmann area 44, in white in left panel) in object-directed and pantomimed actions. The dark gray bands correspond to task periods, whereas the white bands correspond to resting periods. Greater activity is observed for grasping (GO) and touching (TO) an object, compared with a pantomime of grasping (GN) and touching (TN).

formed a functional magnetic resonance imaging (fMRI) experiment in which subjects either performed an object-directed action (grasping or touching an object) or simply pantomimed the action without actually interacting with the object. The prediction is that a hand-object interaction should yield greater activity in regions coding goal-oriented behavior. Consistently with the macaque single-cell data, we found that the inferior frontal cortex had this pattern of activity (figure 2.1).

The early studies on grasping, however, had a conceptual limitation. Even though continuity is important, so that it makes sense to see some features of the action recognition system of the macaque in the human brain, one must also factor in the changes that the evolutionary process might have produced. Thus, to keep focusing on grasping seemed to us a mistake. Imitation seemed a much more promising paradigm to use. In fact, the action recognition system of the macaque has the property of being active both when the monkey performs an action and when it observes an action. These neural properties make this system an ideal candidate for being involved in or at least facilitating imitation. It is true that the imitative abilities of monkeys are limited, but even if one wants to apply the most stringent definition of imitation and thus conclude that monkeys

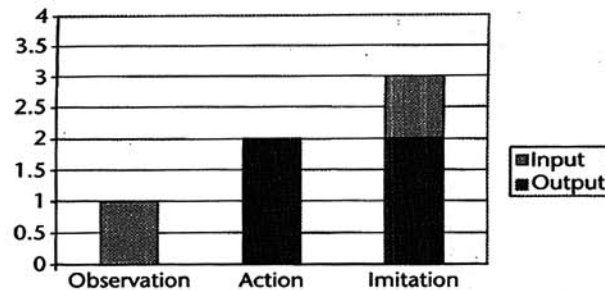


Figure 2.2

The predicted activity for mirror areas during imitation is approximately the sum of the activity observed during observation of an action and during execution of an action.

do not imitate at all, one can also conceivably argue that the action recognition system made monkeys "imitation-ready." Thus it is plausible to predict an involvement of this system in imitation. The way we conceptualized it is captured in figure 2.2. The idea behind this conception is simply that during imitation there is both observation and execution of an action. Thus, one can predict that areas endowed with mirror properties would show an activity pattern similar to the one graphed in the figure, with activity during imitation corresponding roughly to the sum of the activity during observation and execution of action. With the use of fMRI, we found two areas with these properties (Iacoboni et al., 1999). The first area was located in the *pars opercularis* of the inferior frontal gyrus, in inferior frontal cortex, and the second one was located rostrally in the posterior parietal cortex. Thus there was a convincing anatomical correspondence between the areas identified in the human brain as having mirror properties and the macaque mirror areas.

We initially proposed some sort of "division of labor" between the frontal and the posterior parietal mirror areas, so that frontal mirror areas would code the goal of the imitated action and the posterior parietal mirror areas would code somatosensory information relevant to the imitated action. This division of labor was based on considerations inspired by single-cell (Sakata et al., 1973; Mountcastle et al., 1975; Kalaska et al., 1983; Lacquaniti et al., 1995) and neuroimaging data (Decety et al., 1997; Grèzes et al., 1998). Empirical support for this proposed division of labor has been provided recently by an imaging study from our group. The study shows a modulation of activity in inferior frontal mirror areas during imitation of

goal-oriented action, with greater activity during goal-oriented imitation than nongoal-oriented imitation (Koski et al., 2002).

To go back to the first experiment on imitation (Iacoboni et al., 1999), the third region identified by single-cell studies in the macaque as relevant to action recognition, STS, demonstrated a somewhat unexpected pattern of activity. As expected, there was greater activity in STS for action observation than for control visual tasks and for imitation compared with control motor tasks. However, there was also greater activity in STS for imitation than for action observation. This was a somewhat unexpected finding because the observed action was the same during imitation and during action observation. If STS simply encodes the visual description of actions, its activity should be the same during imitation and action observation. Two possible explanations of this finding are as follows: First, the increased activity during imitation may simply reflect increased attention to the visual stimulus because the subjects are supposed to imitate it. Alternatively, the increased STS activity may be due to efferent copies of motor commands originating from the frontoparietal mirror areas. These efferent copies would allow a prediction of the sensory consequences of the planned imitative action that would be compared with the description of the observed action provided by STS. If a good match is obtained, then the planned imitative action can be performed.

To test these contrasting hypotheses, we performed a second fMRI study of imitation in which the subjects were asked to imitate in two different configurations, *specular* (as in a mirror) and *anatomical*. During specular imitation, the subjects moved their right hands to imitate a left-hand action. During anatomical imitation, the subjects moved their right hands to imitate a right-hand action. It turns out that the specular form of imitation is the most common or spontaneous form of imitation early in human development (Wapner & Cirillo, 1968), and it tends to yield a better performance also in adults (Ishikura & Inomata, 1995). Thus we predicted that the specular form of imitation should produce greater activity in mirror areas if they are critical cortical areas for imitation. In fact, in the *pars opercularis* of the inferior frontal gyrus (the human homologue of F5), we found a reliably greater activity for specular imitation than for anatomical imitation (Koski et al., 2003).

This differential activity in the inferior frontal cortex during the two forms of imitation allowed us to test the two contrasting hypotheses about STS, the *attentional* versus the *predictive* hypothesis. If the increased activity in STS during imitation is due to attentional factors, the two forms of imitation, specular and anatomical, should yield similarly increased STS

activity compared with action observation. The anatomical form of imitation, being less natural than the specular one, might yield greater activity as a result of increased attentional demands. In contrast, if the increased activity in STS is due to efferent copies of motor commands with predictive value allowing control of the imitative output, then the STS activity should be similar to the activity in the mirror inferior frontal area, with greater activity for specular imitation than for anatomical imitation. Also, given that efferent copies of motor commands are not produced during action observation, STS activity should not be greater during observation without imitation of a left-hand action than during observation without imitation of a right-hand action. The results shown in figure 2.3 support the predictive hypothesis, with greater activity during specular imitation than

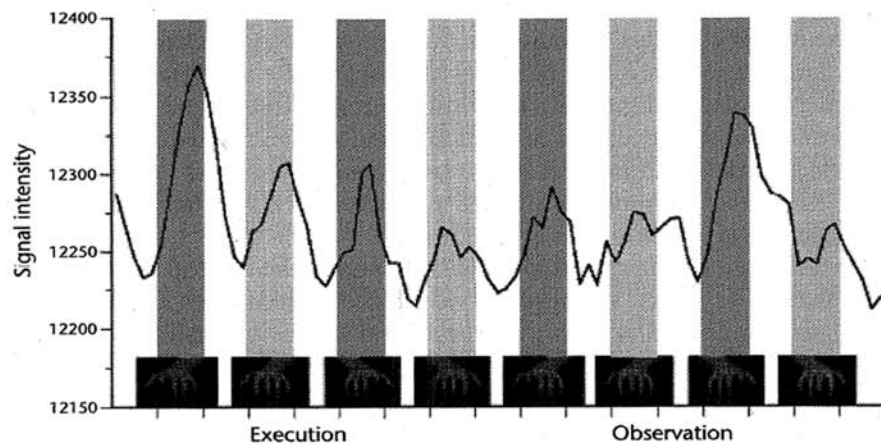


Figure 2.3

Time series of STS activity during imitation and observation of hand movements. The dark bands correspond to task periods and the white bands correspond to resting periods. From left to right, the first four tasks are motor tasks, the last four are observation-only tasks. The hands with the finger lifted up represent task periods when the subjects were shown finger movements. The hands with the cross on the finger represent task periods when the subjects were shown a static hand with a cue on one of the fingers. Thus, the first and third task periods here correspond to *mirror* and *anatomical* imitation, respectively. There is reliably greater activity in STS for *mirror* than for *anatomical* imitation. With regard to observation tasks, observed left-hand actions do not yield greater STS activity, suggesting that the effect observed during imitation is due to efferent copies of motor commands. (Reprinted from Iacoboni et al., 2001.)

during anatomical imitation, but not during left-hand versus right-hand action observation (Iacoboni et al., 2001). This suggests that the increased activity in STS during imitation is due to efferent copies of motor commands that originate in frontoparietal mirror areas and are sent back to STS for monitoring purposes.

To summarize, the information flow within the three areas (superior temporal, posterior parietal, inferior frontal) that form what we call the *minimal neural architecture* for imitation would be as follows:

- The superior temporal cortex provides posterior parietal mirror neurons with a visual description of the observed action to be imitated.¹
- The posterior parietal mirror neurons provide additional somatosensory information regarding the action to be imitated and this information is sent to inferior frontal mirror neurons.
- Inferior frontal mirror neurons code the goal of the action to be imitated.
- Efferent copies of motor commands providing the predicted sensory consequences of the planned imitative actions are sent back to STS.²
- A matching process occurs in STS between the visual description of the action and the predicted sensory consequences of the planned imitative actions. If there is a good match, the imitative action is initiated; if there is a large error signal, the imitative motor plan is corrected until convergence is reached between the superior temporal description of the action and the description of the sensory consequences of the planned action.

This model predicts two things. First of all, the role of STS is extremely important when the action to be imitated is a novel action that is not in the motor repertoire of the imitator. Second, if there is such a robust shuffling of information between STS and frontoparietal mirror areas with regard to imitation, then the body maps of these cortices should have similar organizational principles. The first prediction has never been tested, as far as I know. The second prediction has been tested by us with fMRI. Given that at a premotor and posterior parietal level it is possible to observe a somatotopic representation of body parts, we predicted the topography of body parts in STS. Several laboratories had previously reported human STS

1. A favored information flow from superior temporal to posterior parietal cortex rather than to inferior frontal cortex is postulated on the basis of more robust projections from superior temporal to posterior parietal cortex compared with the inferior frontal cortex.
2. Elsewhere we called these efferent copies reafferent (Iacoboni et al., 2001) to emphasize the information flow going from STS to frontoparietal mirror areas and back to STS.

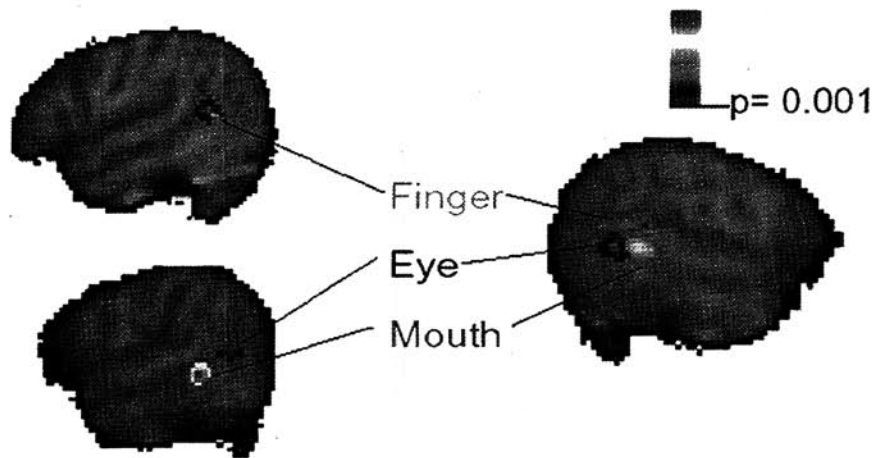


Figure 2.4

Topography of representation of body parts in human STS.

responses to biological motion, but owing to differences in methodology, it was difficult to extrapolate a map of body parts in STS from the published data. We thus performed another fMRI experiment in which the subjects observed hand actions, mouth movements, and eye movements (Dubeau et al., 2001). An orderly topography of body parts was observed in STS, as shown in figure 2.4. Such a topography has been subsequently confirmed by independent observations in our laboratory in separate studies using faces or hands as visual stimuli.

2.2.3 Functional Properties of the Minimal Neural Architecture for Imitation

The temporo-parieto-frontal circuit we delineated in ch. 2.2.2 seems to have relatively well-defined functional properties. Combining information from single-unit and brain imaging data, we have described a plausible information-processing flow that goes from STS to frontoparietal mirror areas, back to STS and then back to frontoparietal areas for the initiation of the imitative movement. In this section we extend the discussion on the functional properties of this model.

The functional properties of the minimal neural architecture for imitation can be mapped onto the functional architecture of modular pairs of forward and inverse models, a computational architecture developed in the motor control literature (Wolpert & Kawato, 1998; Kawato, 1999). Inverse

models are important for motor control, whereas forward models are important for motor learning. An inverse model retrieves the motor plan necessary to reach a desired sensory state. The input of the inverse model is the desired sensory state and its output is the motor plan necessary to reach that state. In our case, the imitator desires to imitate the action of the actor, and an inverse model is created by STS inputting the visual description of the observed action into frontoparietal mirror areas that produce the output of the inverse model, the motor command necessary to imitate the actor. An efferent copy of the motor command (from frontoparietal mirror areas) is fed into STS to create a forward model that allows prediction of the sensory consequences of the planned imitative action. A forward model is a mimic (ironically so, this is imitation after all ...) of the motor system, and if the prediction is confirmed by reafferent feedback, then the pair of forward and inverse models is reinforced by a "responsibility signal" (Haruno et al., 2001) that assigns "high" responsibility for imitating a given action to that specific forward-inverse model pair.

Several questions are left unanswered by this scenario. For instance, how does the pairing of STS input and frontoparietal output that determines inverse modeling occur?³ Remember that STS neurons are driven by hand-object interactions and show visual invariance; that is, under widely different visual circumstances the same neuron will fire at the sight of the same kind of action. The F5 neurons also respond to the sight of a relatively large class of actions under different visual conditions. With regard to their motor properties, however, these neurons are often tuned to a specific action. Thus it is possible that the inverse-forward model pairing is initially facilitated because a large variety of visual stimuli nonselectively activate STS and F5. This nonselective visual activation, however, can be efficiently mapped only onto a few specific motor outputs coded in F5. When an inefficient motor output is selected, it will generate a large error signal in its forward model and will be assigned "low" responsibility for that specific visual input. When an efficient motor output is finally selected, a small error signal will be generated, and when the prediction of the forward model is confirmed by reafferent feedback, the pair of forward and inverse models will be given high responsibility for that action.

3. The same question is addressed by Cecilia Heyes in vol. 1, ch. 6, p. 158. She calls it the "correspondence problem." The solution that she proposes addresses a level of explanation somewhat different from the one we are addressing here. However, what she proposes is not only very plausible, but also is compatible with the functional architecture presented here.

Another question that is worth posing is the following. Within each region (STS, human inferior frontal or macaque F5, human posterior parietal or macaque PF), are the areas receiving inputs and producing outputs the same or different? In other words, how does the functional architecture of inverse and forward models map onto neural structures and above all onto mirror neurons? We recently performed a meta-analysis of some fifty fMRI datasets obtained from normal volunteers performing imitation and observation of hand actions (Molnar-Szakacs et al., 2002). We have observed that in the dorsal sector of pars opercularis of the inferior frontal gyrus (the human homologue of F5) there is an area that is active during both action observation and imitation, whereas in the ventral sector of pars opercularis of the inferior frontal gyrus, there is an area that is active during imitation but not during action observation.

One way of looking at these findings is that the dorsal sector represents the mirror sector of pars opercularis, whereas the ventral sector is simply a premotor area. Another way of looking at them, however, is that the dorsal sector of pars opercularis receives the STS input of the visual description of the observed action and produces the motor plan that forms the output of the inverse model, whereas the ventral sector of pars opercularis produces the efferent copy of the motor command to be sent back to STS and used by the forward model to predict the sensory consequences of the imitative action.

In STS, the meta-analysis that we performed shows that the areas active during imitation and action observation overlap completely. This suggests that the same STS area produces the visual description of the action used as input by the inverse model and receives the efferent copy of the motor command used by the forward model.

2.3 A Way to Language

2.3.1 The Essential Role of Broca's Area in Imitation of Finger Movements

Some neuroanatomical considerations suggest that area F5 of the macaque brain is the evolutionary precursor of Brodmann area 44 (BA 44) of the human brain (von Bonin & Bailey, 1947; Petrides & Pandya, 1994; Rizzolatti & Arbib, 1998). Brodmann area 44 is a cytoarchitectonic area that probabilistically maps onto pars opercularis of the inferior frontal cortex (Mazziotta et al., 2001a,b), an area that we have seen is strongly implicated in imitation and belongs to what we call the minimal neural architecture for imitation. BA 44 is part of Broca's area (some authors assign Broca's area

to BA 44 only), the most important cortical region for language processing in the human brain. The simultaneous involvement of BA 44 in language and imitation, and the evolutionary anatomical considerations that I address later, suggest functional links between imitation and language. A discussion of these possible links is the focus of this section.

In the macaque frontal lobe, there is a major sulcus that divides the anterior granular prefrontal cortex from the posterior agranular motor and premotor cortex. This sulcus is called the arcuate sulcus since its shape resembles an arc. It has been suggested that the evolutionary process has transformed the dorsal sector of the arcuate sulcus of the macaque brain into the superior frontal sulcus of the human brain. The same process would have transformed the ventral sector of the arcuate sulcus into the inferior frontal sulcus of the human brain. Area F5 is ventral to the arcuate sulcus, and its human homologue would also be located ventrally with respect to the inferior frontal sulcus. Thus one would expect to find the human homologue of area F5 in the inferior frontal gyrus. And one would most likely find it relatively posteriorly located, sitting just in front of the premotor cortex represented in the precentral gyrus, given that F5 is a rostral premotor area. If one then considers the cytoarchitecture of F5, which is devoid of a granular layer, one would expect the human homologue of F5 to be an agranular cortical area.

If one considers all these points, the most likely candidate as a human homologue of F5 is BA 44 (Rizzolatti & Arbib, 1998; Geyer et al., 2000). In fact, BA 44 is the rostralmost agranular cortical field in the inferior frontal gyrus and is located right anteriorly to the ventralmost sector of the precentral gyrus. If BA 44 were the human homologue of macaque F5, then one would expect that from a physiological standpoint these two areas should share some features. In terms of motor representation of body parts, F5 in the macaque contains a representation for hand movement and one for mouth movements. Thus one would expect motor representation for the hand and the mouth in BA 44 also. In fact, several imaging studies have reported activation in BA 44 for motor tasks that engage the hand (Krams et al., 1998; Binkofski et al., 1999b; Iacoboni et al., 1999; Ehrsson et al., 2000) and the mouth (Fox et al., 2001). This also means that in an imaging experiment on, say, imitation of foot movements, one should not expect to observe activation of BA 44 if this activation reflects the motor aspect of BA 44 and not its linguistic (supposedly disembodied) one.

The activation studies that we have performed on imitation of hand movements have demonstrated that a sector of pars opercularis is activated during imitation and observation of hand actions. Its activity is also

modulated by the type of imitation (specular versus anatomical) and by goal-oriented imitation. Our meta-analysis of hand imitation (Molnar-Szakacs et al., 2002) points to the dorsal sector of pars opercularis as the sector of Broca's area with these characteristics. Also, a meta-analysis of language tasks (Chein et al., 2002) suggests that the dorsal sector of pars opercularis is a critical language region. The convergence of the empirical data is impressive and suggests shared neural structures for imitation and language. It could be objected, however, that the activation observed in Broca's area during imitation is simply due to some kind of silent and perhaps unconscious verbalization (Heyes, 2001a). This objection does not really explain why there should be more verbalization in some imitative conditions but not in others. It also does not explain why the observation of hand movements activates Broca's area but the observation of foot movements does not. At any rate, the silent verbalization hypothesis cannot be dismissed too lightly.

To test whether the involvement of Broca's area in imitation is due to silent verbalization or to mirror activity for hand or mouth movements, we used repetitive transcranial magnetic stimulation (rTMS). This technique allows one to create a kind of temporary, transient lesion in the brain area stimulated (Walsh & Cowey, 2000). We stimulated the pars opercularis of the inferior frontal gyrus during imitation of hand actions and during a control visuomotor task. We also stimulated a control site. If activation of Broca's area in imaging studies of imitation is due only to an epiphenomenal silent verbalization, then producing a transient lesion in pars opercularis should not affect the imitative performance. If, in contrast, Broca's area is essential to imitation, then producing a transient lesion in pars opercularis should have an effect on imitation. And if this effect reflects impairment in functional processes that are specific to imitation and to BA 44, then performance in the control task should be unaffected by stimulation of pars opercularis, and performance in the imitation tasks should be unaffected by stimulation of the control site. The results we obtained in a recent rTMS study performed in our laboratory (Heiser et al., 2003) are consistent with the hypothesis of an essential role for Broca's area, namely BA 44, in imitation (figure 2.5).

The possibility remains that the effect we observed is language mediated, so that one has to name a finger movement in order to imitate it. However, this is quite unlikely if one looks at human development. In fact, infants can imitate much earlier than they can talk (Meltzoff & Moore, 1977). If Broca's area has an essential role in imitation, then it must be concluded that this area is not exclusively dedicated to language processing. It also

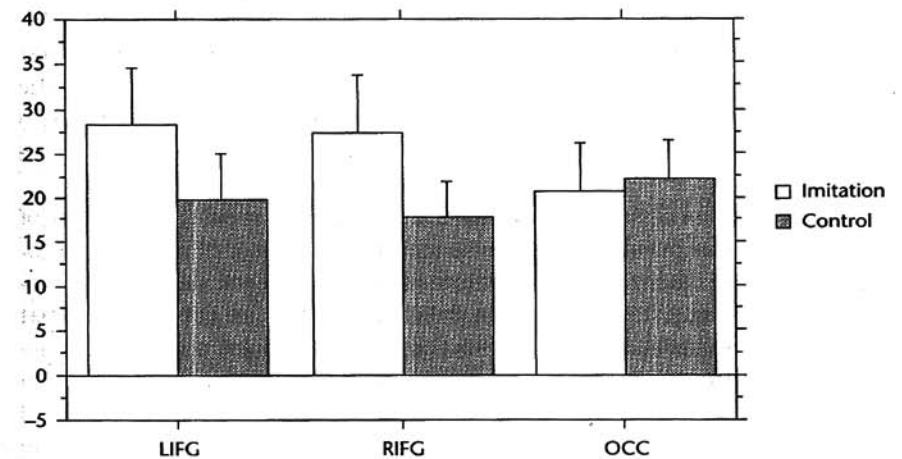


Figure 2.5

Percent errors in imitation (white bars) and control task (gray bars). LIFG, left inferior frontal gyrus (pars opercularis); RIFG, right inferior frontal gyrus (pars opercularis); OCC, control site in occipital cortex. There is a reliable increase in the error rate for rTMS over the left and right pars opercularis during imitation only.

suggests an evolutionary continuity between action recognition, imitation, and language.

2.3.2 Warping Chimp Brains onto Human Brains

The three areas that form what we call the minimal neural architecture for imitation are all located around a major sulcus of the human brain, the Sylvian fissure. The cortex around the Sylvian fissure is called perisylvian and it is known to be extremely important for language. It is possible that the circuit for action recognition in the monkey has evolved to support imitation and subsequently language in humans. From a relatively simple neural mechanism of matching observation and execution of an action (mirror neurons), more complex functional properties were built and more complex behaviors were supported. If this hypothetical scenario has some plausibility, one should observe across species morphometric changes localized around the Sylvian fissure.

The laboratory of Karl Zilles has warped magnetic resonance images of chimpanzees' brains into those of human brains. The process of warping the brain of one species into that of another can be quantified by mapping the local field deformations determined by the warping algorithm. By

doing so, one can obtain a map of the local changes caused by the evolutionary process in the primate brain. What emerges is that the greatest changes can be observed around the Sylvian fissure, thus supporting the hypothesis that the action recognition system of the macaque is the evolutionary precursor of the neural systems associated with language in the human brain.

From a functional standpoint, the mirror system meets the criterion of the “parity assumption” between a sender and a receiver of a message proposed by Alvin Liberman and his motor theory of speech perception (Liberman & Mattingly, 1985; Liberman & Whalen, 2000). In this theory, what counts for the sender must count for the receiver. Hence, a common code for language perception and language production is necessary. Evidence in favor of such a code in the domain of language has been recently provided by a TMS study in which subjects listened to speech. It was found that the listeners’ tongue muscles were much more excited by a single TMS pulse delivered over the motor cortex when they were listening to words that required strong tongue movements to pronounce (Fadiga et al., 2002). This evidence is clearly compatible with the motor theory of speech perception, according to which it is the activation of the articulatory motor gesture that enables the perception of the speech sounds.

Another empirical link between speech perception and the action recognition system of the macaque has been recently provided by single-unit data. It has been shown that mirror neurons in the macaque area F5 fire, not only at the sight of an action, but also at the sound of an action (i.e., breaking a peanut) in the dark (Kohler et al., 2002). These data suggest two things. First, mirror neurons have the auditory access necessary to implement speech perception. Second, they enable a multimodal representation of action that is not linked to the visual channel only. This may facilitate the learning of speech sounds via imitation.

However, the question that is typically raised here is, How does one go from a relatively simple action recognition system to the complex symbolic levels reached by human language? In the next section, rather than trying to answer the question, I look at the plausibility of the question itself and discuss language, not in the abstract forms often studied by mainstream linguistics, but in its daily, embodied form: conversation.

2.3.3 What We Talk about When We Talk about Language

Traditionally, linguists have approached language by trying to extrapolate the most formal and abstract aspects of it. As shown in figure 2.6 (left side), one can form sentences following what has been called a tree structure. The

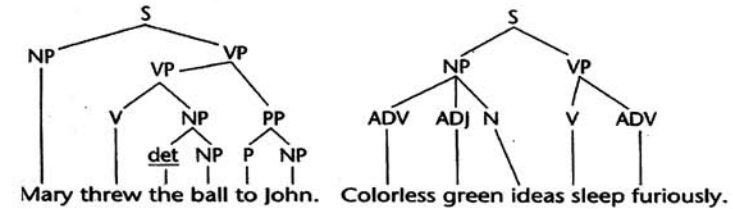


Figure 2.6

main idea behind it is that certain lexemes “govern” or “bind” some other lexemes. Thus, every grammatical sentence (S, top of the tree) consists at least of a noun phrase (NP) and a verb phrase (VP). NP and VP in turn break down into various “branches.” An NP may consist of a determiner and an adjective and a noun, as for instance in the case “the funny guy.” A VP may consist of a verb (threw) and an NP (the ball) and a prepositional phrase (to John) that can be further decomposed.

The benchmark test of this approach is shown in figure 2.6 (right side). The sentence is completely nonsensical and doesn’t refer to real things, yet it still follows grammatical rules. This has been taken as evidence in favor of a universal grammar that can be processed similarly by all brains (Chomsky, 1981, 1986a,b, 1990, 1997, 1999). So the question here would be, How can such a formalized structure emerge from a relatively primitive action recognition system? A type of answer (very vague, admittedly) to this question that is provided by others elsewhere (Rizzolatti & Arbib, 1998) is that “gestures may be a primitive form of grammar.” The problem with both question and answer is that they accept a view of language as a phenomenon that can be essentially reduced to formal constructs such as grammar.

The real question to ask is, How do people talk? Figure 2.7 shows a transcription of typical naturally occurring everyday speech. As can be seen, not only do violations of grammar (in italics) occur throughout the speech, but most importantly, this segment of speech is full of phenomena (indicated by question marks that show sound stretches, hesitation markers, false starts, self-repairs, prosodic emphasis, and manipulation of timing and word rhythm) that are meaningful to both speaker and hearer. These phenomena are not even part of what is studied by traditional linguists. However, Conversation Analysis (CA) has provided a corpus of robust empirical data that describe these phenomena well (Heritage, 1989; Goodwin & Heritage, 1990; Heritage & Roth, 1995; Goodwin, 2000; Goodwin & Goodwin, 2000).

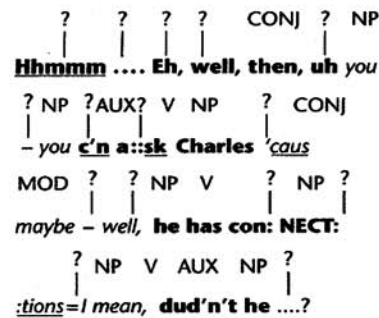


Figure 2.7

A salient feature of typical conversations that is ignored by traditional linguists is turn-taking. The average transition space from one speaker to another is less than 0.2 seconds, and longer pauses are immediately perceived as violations of temporal norms, even by young children. What enables such fast transitions is the hearer's tracking through the stream of the other's speech the appearance of fairly invariant, transiently appearing opportunities for taking a turn—which are not exclusively grammatical but rhythmic and pitch contour-intonational as well. CA data show that the fast transition that typically occurs between sentences is so familiar to both parties as a legitimate point for the other speaker to come in that elaborate strategies in rhythm, intonation, and even grammar have to be adopted by the original speaker in order to obtain a multisentence turn in conversation.

What comes out of CA's acknowledgment of language as an embodied practice is that such motoric processes as eye-gaze, body torque, rhythm attunement, and simultaneous gestures are part of a social interaction (rather than a "software program" as classic cognitivism advocates) that is critically dependent on the motor system's facility for temporal orientation and organization of sequence and, I propose, is also dependent on (and plausibly even deriving from) the action recognition or mirror system.

Thus, rather than trying to mimic a traditional grammatical structure in the domain of manual communication (the approach of Rizzolatti and Arbib, 1998), what I advocate here is to consider carefully the incontrovertibly motor elements that are at work in conversation and that human beings must still rely on to choreograph their interactions with each other. The processing of all those motor elements that give meaning to conversations requires a fast functional architecture that is not dissimilar to the

one needed in motor control. Here is where forward and inverse modeling and the action recognition system, in which mirror neurons can coordinate activity between individuals with a simple matching mechanism, have a real computational advantage over classical cognitive architectures.

2.4 Feeling the Emotions of Others

2.4.1 Empathy and Imitation

Empathy allows the sharing of experiences, needs, and goals across individuals, thus playing a fundamental role in social cognition. The functional aspects and corresponding neural mechanisms of empathy, however, are poorly understood. When Theodore Lipps introduced the concept of empathy (*Einfühlung*), he theorized a critical role for the mechanism of *inner imitation* of the actions of others in generating empathy (as cited in Gallese, 2001). In support of Lipps's idea, empathic individuals exhibit non-conscious mimicry of the postures, mannerisms, and facial expressions of others (the *chameleon effect*) to a greater extent than nonempathic individuals (Chartrand & Bargh, 1999). Thus empathy may occur via a mechanism for representing action that modulates and shapes our understanding of the emotional states of other individuals.

Separate neural systems for emotions and action representation, however, do exist in the primate brain. The limbic system is critical for processing emotion and behavior, and the temporo-parieto-frontal circuit I described earlier (the minimal neural architecture for imitation) is critical for representing action. Anatomical data suggest that a sector of the insular lobe, the dysgranular field, is connected with the limbic system as well as with posterior parietal, inferior frontal, and superior temporal cortex (Augustine, 1996). This connectivity pattern makes the insula a candidate for relaying information about the representation of actions to limbic areas that process emotional content.

2.4.2 The Minimal Neural Architecture for Imitation and the Limbic System: A Role for the Insular Lobe

To test this hypothesis, we performed two experiments—a brain imaging experiment with normal volunteers and a neuropsychological study with both neurological patients and normal controls. In the brain imaging study (Carr et al., 2003) we used fMRI while the subjects were either observing or imitating emotional facial expressions. A modulation of the action representation circuit onto limbic areas via the insula predicts greater activity in the whole network during imitation, compared with observation of

emotion. In fact, mirror areas would be more active during imitation than during observation because of the simultaneous encoding of the sensory input and planning of the motor output. Within mirror areas, the inferior frontal cortex seems particularly important here, given that an understanding of goals is an important component of empathy. The insula would be more active during imitation because its role as a relay would increase, compared with mere observation. Finally, limbic areas would also increase their activity because of the increased motor activity. Moreover, if mediation by representations of action is really critical to empathy and the understanding of the emotions of others, then even the mere observation of emotional facial expressions should activate brain regions of motor significance. Thus observation and imitation of emotions should yield substantially similar patterns of activated brain areas, with greater activity in premotor cortex, especially inferior frontal cortex, and in superior temporal cortex, insula, and limbic areas during imitation.

The results of the fMRI study confirmed our hypothesis. There was a substantially similar network of activated areas for both imitation and observation of emotion. Among the areas activated during both imitation and observation, the premotor face area, the dorsal sector of pars opercularis of the inferior frontal gyrus, the superior temporal sulcus, the insula, and the amygdala had greater activity during imitation than during observation of emotion.

The peak of activation in primary motor cortex during imitation of facial emotional expressions that we observed in our study corresponds extremely well with the location of the primary motor mouth area as determined by a meta-analysis of published PET studies, by a meta-analysis of original data in thirty subjects studied with PET, and by a consensus probabilistic description of the location of the primary motor mouth area obtained by merging the results of the two previously described meta-analyses (Fox et al., 2001). This confirms the robustness and reliability of the data, in spite of the presence of facial motion during imitation. This is because, even though motion artifacts were present at the individual level, the group analysis got rid of them since each subject had different kinds of motion artifacts and they were thus eliminated when all the data were considered. In keeping with this, the data also clearly show peaks of activity in the pre-SMA face area and the face area of the posterior portion of the rostral cingulate zone (RCZp) that correspond extremely well with the pre-SMA and RCZp face locations as determined by a separate meta-analysis of PET studies that focused on motor areas in the medial wall of the frontal lobe (Picard & Strick, 1996).

Thus, our dataset clearly represents the first fMRI demonstration of human primary motor and rostral cingulate face areas. With regard to premotor regions, the peaks that we observed correspond well with premotor mouth peaks as described by studies in which action was observed. In fact, robust premotor responses during the observation of facial emotional expressions were found, which is in line with the hypothesis that action representation mediates the recognition of emotions in others even during simple observation.

The activity in pars opercularis shows two separate foci during imitation, a ventral and a dorsal peak, but only the dorsal peak remained activated, albeit at significantly lower intensity, during observation of emotion. This pattern, with very similar peaks of activation, was also observed in our fMRI meta-analysis of imitation and observation of hand actions previously described (Molnar-Szakacs et al., 2002). In the monkey, F5 neurons coding arm and mouth movements are not spatially segregated, and our human imaging data are consistent with this observation. Furthermore, the imaging data on imitating facial emotion converge with the data on hand imitation in suggesting that the mirror sector of the human inferior frontal gyrus is located in the dorsal part of pars opercularis.

The anterior sector of the insula was active during both imitation and observation of emotion, but more so during imitation, fulfilling one of the predictions of the hypothesis that action representation mediates empathy. This is in line with two kinds of evidence available on this sector of the insular lobe. First, the anterior insula seems to receive slow-conducting unmyelinated fibers that respond to a light, caresslike touch and may be important for emotional and affiliative behavior between individuals (Olausson et al., 2002). Second, imaging data suggest that the anterior insular sector is important for the monitoring of agency (Farrer & Frith, 2002), the sense of ownership of actions, which is a fundamental aspect of action representation. This confirms a strong input into the anterior insular sector from areas of motor significance.

The increased activity in the amygdala during imitation compared with observation of emotional facial expression reflects the modulatory role of the action representation circuit in limbic activity. It has been long hypothesized (Darwin was the first in 1871; Ekman, 1973, 1999; Buck, 1980) that facial muscular activity influences people's affective responses. This is the first demonstration, however, that activity in the amygdala, a critical structure in emotional behaviors and in the recognition of facial emotional expressions of others, increases while subjects imitate the facial emotional expressions of others, compared with mere observation.

To further test the hypothesized mediation of limbic activity by action representation via the insula, we also studied the ability to recognize emotions in two groups of ischemic stroke patients and in a control group. We studied patients with insular lesions only and compared them with patients who had lesions outside the insula. If empathy occurs via modulation, implemented in the insula, of limbic activity by action representation, then an insular lesion should impair such a mechanism. This predicts a greater deficit in recognizing emotion in patients with insular lesions than in patients with lesions outside the insula. This is because the insula would be an obligated path through which the modulation of action representation networks must be channeled to reach limbic structures such as the amygdala. This prediction was confirmed by the empirical data. Although both groups of patients were significantly impaired compared with normals, the deficit in recognition of emotion was significantly greater in insular patients than in stroke patients with lesions outside the insula. The deficit in recognition of emotion in noninsular patients may be due either to nonspecific effects of the lesions or to some effects on the action representation network produced by the lesions. The additional evidence from the data obtained in neurological patients is extremely important because activation studies cannot fully determine if an activated area is essential to the task or behavior studied. Recently, repetitive TMS has been used to create transient lesions in normal subjects to test how essential a given activated area is, as in our rTMS study on imitation in Broca's area described earlier. Repetitive TMS, however, cannot reach deep structures such as the insula. Hence the clinical data are essential here.

Some preliminary neurological evidence in support of both the anatomical and functional proposal described here has also been described in a patient with a subcortical lesion encompassing the anterior sector of the insula. This patient had both the inability to detect disgust as presented in many different ways, from nonverbal emotional sounds to prosody to facial expressions, and the inability to experience disgust himself (Calder et al., 2000).

All in all, we understand the feelings of others via a mechanism of action representation that shapes emotional content, such that our empathic resonance is grounded in the experience of our acting body and the emotions associated with specific movements. As Lipps noted, when I observe a circus performer on a hanging wire, I feel I am inside him (as cited in Gallese, 2001). In order to empathize, we rely on mediation by the representation of the actions associated with the emotions we are witnessing and on a brain network that includes structures supporting communication

between action representation circuits and circuits dedicated to emotional processing.⁴

2.5 Conclusion

The temporo-parieto-frontal circuit described here is the first available neural model of imitation in primates. The anatomical location and the functional properties that this circuit exhibits confirm a key role for imitation in learning and communication. This may not appear as big news to behavioral scientists who have studied imitation for years and have observed how imitative abilities are tied to social learning. The novelty, however, resides in the fact that our research program shows how the functional properties of a relatively well-developed large-scale neural circuit can now inform us about the functional characteristics of behavioral domains that remained for a long time impenetrable to a neuroscientific investigation. This information, in turn, can be used to test more general questions in the behavioral sciences; for instance, the innateness or learnability of some functions, or the approach human beings take to mentalizing.⁵

Acknowledgments

This work was supported in part by the Brain Mapping Medical Research Organization, the Brain Mapping Support Foundation, the Pierson-Lovelace Foundation, the Ahmanson Foundation, the Tamkin Foundation, the Jennifer Jones-Simon Foundation, the Capital Group Companies Charitable Foundation, the Robson Family, the Northstar Fund, and grants from the National Center for Research Resources (RR12169 and RR08655) and the U.S. National Science Foundation (REC-0107077). I wish to thank Don Favareau for his contribution to this chapter. Without his data, knowledge, and enthusiasm, section 2.3.3 could not possibly have been written.

4. For a discussion relevant to this section, see vol. 2, ch. 13 by Jesse Prinz and the comments by Huesmann in vol. 2, ch. 19.6, p. 386. ED.

5. See comments on this chapter by Arbib, vol. 1, ch. 8.2, p. 200, by Blackmore, ch. 8.3, p. 203. ED.