In S. Hurley & N. Chater (Eds) 2005. Perspectives on Imitation MIT Press

3 "Being Like Me": Self–Other Identity, Mirror Neurons, and Empathy

Vittorio Gallese

3.1 Introduction

We readily ascribe intelligence to other animals while being simultaneously inclined to think that—cognitively speaking—humans "do it better." We are and we feel we are different from other animals, even from our closest relatives among nonhuman primates, the apes. There are indeed many differences between humans and other primates. One of the most crucial is thought to be the capacity to "read" the mind of others, which many ascribe only to humans.

In daily life we are constantly exposed to the actions of other individuals inhabiting our social world. We are not only able to experience their behavior, understand its content, and predict its consequences, we can do more than that; we can also attribute intentions to other individuals. We can immediately recognize whether their behavior is the result of a purposeful and deliberate attitude or the unpredicted consequence of some accidental event that is totally unrelated to their will. As maintained by so-called "folk psychology," we are able to understand the behavior of others in terms of their mental states. This view prefigures a distinction between species that are confined to behavior reading and our species, which makes use of a different level of explanation: mind reading.

However, it is by no means obvious that behavior reading and mind reading constitute two autonomous, encapsulated realms. It is even less obvious that in understanding the intentions of others we employ a cognitive strategy totally unrelated to predicting the consequences of their observed behavior. Whenever we face situations in which exposure to others' behavior requires a response by us, be it active or simply attentive, we seldom engage ourselves in an explicit, deliberate interpretative act. Our understanding of a situation most of the time is immediate, automatic, and almost reflexlike. Therefore it seems preposterous to claim that our capacity

to reflect on the real intentions determining others' behavior is all there is to understanding it.

Mind reading, whatever it might be, is at best only one part of our mental space. This space is multidimensional; it is as many-sided as the dimensions that characterize our mental life and as the many possible ways to live our lives and to look at them. We can put ourselves on a scale and check our body weight. Or we can think about what someone else shouldn't have thought about us. In both instances we do not experience any identity shift. We do not feel *different* when we are checking our body weight and when we entertain counterfactual third-person metarepresentations. This is quite rightly so, in that what does change is not the individual organism. What changes is the type of *relational specification* by which each organism (a biological system) engages itself during the various possible kinds of interaction with the world outside. Relational specifications constitute the almost infinite levels at which we may decide to *act upon* the world. And there are almost infinite levels at which *others* may do the same. We can take a swim, plant a tree, get a doctoral degree, or think about Ulysses, while simultaneously knowing in an implicit and unmediated way that others do the same and think the same, or that they do not. All these levels of interaction, when ascribed to others, pertain to different beings, different persons whom, nevertheless, we feel, recognize, and represent as similar to us.

Beside—and likely before—the ascription of any intentional content to others, we entertain a series of implicit certitudes about the content-bearing individuals we are confronted with. These implicit certitudes constitute the intersubjective relation and concern the sense of oneness, of identification with the other that makes sit possible to ascribe *any* content, whatever it might be, to the individual we are interacting with.

We could certainly hold a solipsistic view and claim that just because all individuals are the same, in defining cognition we should not waste our time with speculations on the relevance of others' minds. Solipsism recommends instead a focus on the *single* individual's mind. This should secure enough knowledge to define what a mind is and how it works. Following this perspective, the mechanisms enabling the epistemic relations between the rational agent and the world are of no relevance for the determination of representational content and for the understanding of what that content is and what it stands for (see Fodor, 1998).

But I will not adopt the solipsistic view. In this chapter I analyze from a neuroscientific perspective the constituents of the implicit certitudes enabling intersubjective relations, and what might be the neural mechanisms

underpinning them. *Pace* solipsism, I propose that our cognitive stance toward life is but one expression of the many and diversified modes in which we interact with the world. From the very beginning of our lives, intersubjective relations constitute a major part of our daily interactions with the world. I will posit that intersubjective relations play a *major* and *constitutive* role in shaping our cognitive capacities and in providing the shared database required to establish meaningful bonds with other individuals.

After having identified the peculiar perspective inspiring the present proposal, let us have a closer look at social behavior. The pervasive social habits of primates are most likely the result of a very long evolutionary path in that these habits are patently not peculiar to primates. They are indeed diffused across species spaced as far apart in evolutionary time as humans and ants. Social interactions play different roles according to different modalities in different species. Nevertheless, transverse to and at the basis of all social species and all social cultures, of whatever complexity, is the capacity for identification with the individuals within those species and cultures. When I speak of self–other identity in this context, I mean the identification of the self with another individual as "like me" in some way (which can, but need not, involve mental identification). As humans, we implicitly know that all human beings have four limbs, walk in a certain way, act in peculiar ways, etc. If we share the same culture, we will, for example, all tattoo our body in a peculiar striped fashion, or wear the same school necktie at reunions, or be against the death sentence, etc.

Identity, as we have seen, is articulated on many different levels of complexity. Identity can be subjected to increasingly complex tests in which different species might score differently, but it is nevertheless the membership fee all individuals have to pay in order to self-guarantee the sense of belonging to a larger community of other organisms. Identity is so important within a group of social individuals because it enables them to predict more accurately the consequences of others' future behavior. This capacity in turn contributes to optimizing the employment of cognitive resources by reducing the meaning space to be mapped. Identity contextualizes content by reducing the number of possible information units the brain is required to process.

Several developmental psychology studies have shown that the identity-based capacity to predict others' behavior is a very early endowment of human beings. In infants the establishment of relations with others is accompanied by the registration of behavioral invariance. This in turn translates into the implicit procedural memory of the organism (on this point and for a discussion of the relevant literature, see Stern, 1985). This

experience-driven process of constant remodeling of the system is one of the building blocks of cognitive development, and it capitalizes upon coherence, regularity, and predictability. Self–other identity underlies all these features, henceforth its high social adaptive value.

Anytime we meet someone, we do not just perceive that someone to be, broadly speaking, similar to us. We are implicitly aware of this similarity because we literally embody it. Meltzoff & Brooks (2001) have convincingly suggested that the "like me" analogy between infant and caregiver is *the* starting point for the development of (social) cognition. This analogical process proceeds in a bidirectional way. Infants use the observed behavior of their human companions as a mirror to gain more knowledge about themselves. But the same process also works the other way around; it enables infants to know about the others.

The posited important role of self–other identity relations in determining the cognitive development of our mind provides a strong motive to investigate from a neuroscientific perspective the functional mechanisms (and their neural underpinnings) at the basis of self–other identity. This is the main issue addressed in this chapter. Later on I discuss the neuroscientific results in relation to the notion of empathy, which, after several decades of almost complete oblivion, has forcefully reappeared in the contemporary debate on human cognition. After a brief historical review, I provide an enlarged account of empathy defined by means of a new conceptual tool: the *shared manifold* of intersubjectivity. I conclude by proposing that it is by means of this shared manifold that other human beings can be recognized to be similar to us. This identity relation will bootstrap imitation, interindividual communication, and mind reading.

3.2 "Being Like Me": A Neuroscientific Approach to the Self–Other Identity

One of the major contributions to a new understanding of human social cognition during recent decades has come from research in developmental psychology. As infants, for years we all heavily rely on interactions with our caregivers and with other individuals to learn how to cope with the world. Developmental psychology has provided an enormous amount of data that have literally revolutionized our way of looking at newborns and infants as cognitive agents. These results have shown, among other things, that at the very beginning of our life we almost immediately interact with others by *reproducing* some of their behaviors. The seminal study by Meltzoff and Moore (1977) and the subsequent research field it opened (see

Meltzoff & Moore, 1997; Meltzoff, 2002a,b; and Meltzoff, vol. 2, ch. 1), showed that newborns as young as 18 hours are capable of reproducing mouth and face movements displayed by the adult they are facing. The particular part of their body replies, although not as a mere reflex (see Meltzoff & Moore, 1977, 1994), to movements displayed by the equivalent body part of someone else. More precisely, this means that newborns set into motion, and in the correct way, a part of their body they have no visual access to, but which nevertheless acts to match an observed behavior. To put it very crudely, visual information is transformed into motor information.

This apparently innate mechanism has been labeled active intermodal mapping (AIM; see Meltzoff & Moore, 1997). Intermodal mapping defines a "supramodal act space" (Meltzoff, 2002a), which provides representational frames not limited to any particular mode of interaction, be it visual, auditory, or motor. Modes of interaction as diverse as seeing, hearing, or doing something *must* therefore share some peculiar feature that makes the process of equivalence carried out by AIM possible.

The issue then consists in clarifying the nature of this peculiar feature and the possible underlying mechanisms. My best candidate for a shared feature is the relational character intrinsic to any interaction between a biological system and the environment. Our environment is composed of a variety of lifeless though not refractory forms of matter and a variety of living things, whose peculiar character is more and more discerned by the infant's immature eye. Individuals confront many possible kinds of external objects and, because of their peculiar status as biological systems, are constrained in their modes of interaction. Any interaction requires a control system implementing a control strategy. Interestingly enough, control strategies share with modes of interaction a *relational* character. As modes of interaction, control strategies are essentially relational in that they *model* the interaction between organism and environment, to better control it.

However, a model is a form of representation. This step allows a relation of interdependence, if not even superposition, to be established between control of behavior and the representation to be established (see Gallese, 2000b). This relation holds for both organism–object and organism–organism modes of interaction. This relation is established at the very onset of our life, when no subjective representation can yet be entertained by us, because there is not yet a *conscious subject* of experience. The absence of a subject does not preclude, however, the presence of a primitive *self–other space*, a paradoxical form of intersubjectivity without subjects. The infant

shares this space with lifeless objects as well as with living others, which are internalized by the infant because they are a projection of the control strategies governing the interactions they are part of. Both lifeless objects and living others are represented as the materialization of their implicit objectual character within these interactions. The physical space occupied by inanimate objects and bodies of the adult others is connected to the body of the infant to compose a blended, shared space.

What is the role and fate of this peculiar shared informational space in the course of cognitive development? This issue is worth scrutiny. The shared blended space enables the social bootstrapping of cognitive and affective development. Once the crucial bonds with the world of others are established, this space carries over to the adult conceptual faculty of socially mapping sameness and difference ("I am a *different subject*"). The more mature capacity to segregate the modes of interaction, together with the capacity to carve out of the blended space the subject and the object of the interaction, does not annihilate the shared space.

The shared space provides an incredibly powerful tool for detecting and incorporating coherence, regularity, and predictability in the course of an individual's interactions with his or her environment. The shared space is progressively joined by perspectival spaces defined by the establishment of capacities to distinguish the self from others while self-control is developing. Within each of these perspectival spaces information can be further segregated in discrete channels (visual, somatosensory, etc.), making our perceptual view of the world more finely grained. The concurrent development of language probably contributes to further separating out of single characters or modalities of experience from the original multimodal perceptual world, but the shared intersubjective space does not disappear. It progressively acquires a different role: to provide our self with the capacity simultaneously to entertain self–other identity and difference. Within intersubjective relations, the other is a living oxymoron, being just a *different self*.

My proposal is that the "selfness" quality we readily attribute to others, the inner feeling of "being like me" triggered by our encounter with others, is the result of this preserved blended intersubjective space. Self–other physical and epistemic interactions are shaped and conditioned by the same body and environmental constraints. This common relational character is underpinned at the level of the brain by neural networks that compress the redundant "who did it," "who is it" specifications, and realize a thinner content state, which specifies what kind of interaction or state

is at stake. This thinner content is shared just because, as we have learned from developmental psychology, the shareable characters of experience and action are the earliest constituents of our life.

Before presenting empirical evidence to support my hypothesis, it is necessary to clarify the conditions under which the neuroscientific level of description would appear reasonably apt to support it. The following conditions should do the job;

1. evidence of a neural representational format that can achieve sameness of content in spite of the specific quality of the mode of presentation of its referents;
2. indifference of the representational format to the peculiar perspective spaces from which referents project their content; in other words, indifference to self–other distinctions;
3. persistence of the same representational format into adulthood.

In the next sections I review neuroscientific evidence from our laboratory that appears to be in a good position to satisfy all three conditions.

3.3 Interactions and Their Models

The most rostral sector of the ventral premotor cortex of the macaque monkey controls hand and mouth movements (Rizzolatti et al., 1981, 1988; Kurata & Tanji, 1986; Hepp-Reymond, 1994). This sector, which has specific histochemical and cytoarchitectonic features, has been termed area F5 (Matelli et al., 1985). A fundamental functional property of area F5 is that most of its neurons do not discharge in association with elementary movements, but are active during *actions* such as grasping, tearing, holding, or manipulating objects (Rizzolatti et al., 1988).

What is coded is the relation, in motor terms, between the organism and the external object of the interaction. Furthermore, this relation is of a very special kind: a relation projected to an expected success. A hand reaches for an object, it grasps it, and does things with it. F5 neurons become active only if a particular type of interaction (e.g., hand–object, mouth–object, or both) is executed until the relation leads to a different state of the organism (e.g., to take possession of a piece of food, to throw an object away, to break it, to bring it to the mouth, to bite it). Particularly interesting in this respect are grasping-related neurons that fire any time a monkey *successfully* grasps an object, regardless of the effector employed, be it any of its two hands, the mouth, or both (Rizzolatti et al., 1988; see also Rizzolatti et al., 2000).

The independence between the nature of the effector involved and the end state that the same effector attains constitutes an *abstract* kind of representation. The firing of these neurons instantiates the same content (the new end state the organism will attain), even if it is differently mediated. In accord with information theory, a thinner content state has been reached by compressing redundant information about which effector or which dynamic parameters should be involved in the interaction. This compression process is not cognitive per se. It is just an information compression process. Nevertheless, by employing an intentional language, we could describe this neural mechanism in terms of goal representation (see Rizzolatti, 1988; Gentilucci & Rizzolatti, 1990).

Beyond purely motor neurons, which constitute the overall majority of all F5 neurons, area F5 also contains two classes of visuomotor neurons. Neurons of both classes have motor properties that are indistinguishable from those of the earlier-described purely motor neurons, while they have peculiar visual properties. The first class is made up of neurons that respond to the presentation of objects of particular size and shape in the absence of any detectable action aimed at them, either by a monkey or an experimenter. The monkey sees a particular object and the neuron fires. These neurons have been labeled canonical neurons (Rizzolatti et al., 1988, 2000; Rizzolatti & Fadiga, 1998).

The second category is made up of neurons that discharge when the monkey *observes* an action made by another individual and when it *executes* the same or a similar action. We labeled them mirror neurons (Gallese et al., 1996; Rizzolatti et al., 1996a; see Rizzolatti, vol. 1, ch. 1, and also Rizzolatti et al., 2001).

Let us first have a closer look at canonical neurons. Most grasping actions are executed under visual guidance. A relationship therefore has to be established between the features of objects and the particular motor specifications they might engender *if* the organism is aiming at them. The appearance of a graspable object in the visual space must somehow set in motion the retrieval of the appropriate mode of interaction required by the intended type of hand–object relation. Suppose we discover neurons that not only code for the motor acts they are supposed to control but also respond to the visual features that trigger them. We would then have a representational format for sameness of content (the successful end state of the hand–object interaction) regardless of the referent, be it the effector or the target object.

Indeed, canonical neurons respond to the visual presentation of objects of different sizes and shapes in the absence of any detectable movement

by the monkey (Rizzolatti et al., 1988, 2000; Jeannerod et al., 1995; Murata et al., 1997). Very often a strict congruence has been observed between the type of grip that activated a neuron and the size and shape of the object that triggered the same neuron's response during mere observation of the object. But there is more; in the observation modality, a considerable percentage of neurons display an equally strong response to objects that although differing in shape, nevertheless all "afford" the same type of grip.

A possible interpretation of these findings is that canonical neurons instantiate a *multimodal* representation of organism–object relations. This representation is originally "motor" because it is triggered and driven by motor-control constraints. It is no coincidence that canonical neurons are part of the premotor cortex. However, the representation they instantiate loses its intrinsic motor quality once it blends with the information fed by visual and auditory (see section 3.4) channels. What is represented is not only (or perhaps not anymore) a motor plan; it becomes a multimodal semantic node.

The human brain is not different in this respect. Brain imaging studies in humans have shown an unexpected correlation between categorical perception of tools and the activation of premotor brain sectors (for review, see Martin & Chao, 2001; Malach et al., 2002; see also Gallese, 2003a). The experiments on monkeys described earlier shed light on the neural mechanism as the basis for these results in humans, which further corroborates the hypothesis proposed here.

These results are important because they emphasize that the intentional character, the "aboutness" of the representational format of our mind, is deeply rooted in the essentially relational character of body action, which in turn suggests the essentially *intertwined character* of action, perception, and cognition (see Hurley, 1998; Gallese, 2000b).

Representational content, and thus a *fortiori* conceptual content, cannot be fully explained without considering it as the result of the ongoing modeling process of an organism. The intrinsic need of any organism to control its dynamic interaction with the environment also constrains the way these interactions need to be modeled and hence represented. The same *sensorimotor* circuits that control the ongoing interactions of an organism with its environment also map objects and events in that environment, thus defining and shaping their representational content. Our representation of the world is a model of the world that must incorporate our idiosyncratic way of interacting with it. As will become clearer in the next section, this feature is not unique to organism–object interactions but also applies to interpersonal relations.

3.4 Self–Other Identity and Shared Multimodal Content

Let us return to neurophysiological data on monkeys from our laboratory. As briefly mentioned in section 3.3, the second class of F5 visuomotor neurons is made up of mirror neurons. They discharge both when a monkey makes a specific action and when it observes another individual making a similar action (Gallese et al., 1996; Rizzolatti et al., 1996a).

This evidence demonstrates that in *adult* individuals, both monkeys and humans (see Rizzolatti, vol. 1, ch. 1), a mirror matching neural mechanism can represent content *independently of the self–other distinction*, thus satisfying the last two criteria I posited to be necessary to ground my working hypothesis empirically. The first criterion, namely, sameness of content regardless of how the referents are presented, has not yet been addressed. In a recent study we investigated whether there are neurons in the monkey premotor cortex that discharge when the monkey makes a specific hand action and also when it *hears* the corresponding action-related sounds. The results showed that the monkey premotor cortex contains neurons that discharge when the monkey *executes* an action, *sees*, or just *hears* the same action performed by another agent. We have labeled these neurons audiovisual mirror neurons (Kohler et al., 2001, 2002). They respond to the sound of actions and discriminate between the sounds of different *transitive* manual or oral actions that are compatible with the monkey's natural behavioral repertoire. Audiovisual mirror neurons, however, do not respond to other similarly interesting sounds, such as arousing noises, or monkeys' and other animals' vocalizations. The actions whose sounds evoke the strongest responses when heard also trigger the strongest responses when they are observed or executed. The activity of this neural network does not significantly differ if events in the world, such as noisy actions, are specified at the motor, visual, or auditory level. Such a neural mechanism can represent the end state of the interaction independently from its different modes of presentation by sounds, visual images, or willed, deliberate acts of the body. All modes of presentation of the event are blended within a circumscribed, informationally thinner level of semantic reference.

Furthermore, and most important for our quest for a neural correlate of intersubjective identity, sameness of content is shared with different organisms. This shared semantic content is the product of modeling the observed *behavior* as an *action* with the help of a matching equivalence between what is observed or heard and what is executed.

Mirror neurons, like canonical neurons, instantiate a *multimodal* representation of organism–object relations. In the case of canonical neurons,

these relations imply an interacting actor; thus they typically pertain to an actor-centered frame of reference. The object is relevant for someone who will do things with it, even if only potentially. However, mirror neurons also do something different. They map this multimodal representation across different spaces inhabited by different actors. These spaces are blended within a unified common intersubjective space, which paradoxically does not segregate any subject. This space is *"we" centric*.

It is worth mentioning that in both monkeys and humans, the mirror system has been discovered and studied in adult individuals (see Rizzolatti, vol. 1, ch. 1). This means that in humans, and even more so in monkeys, the shared space coexists with but does not determine self-awareness and self-identity. The shared intentional space underpinned by the mirror matching mechanism is not meant to distinguish the agent from the observer. As organisms we are equipped with plenty of systems, from proprioception to the expectancy created by the inception of any activity, that are able to distinguish the self from the other. Rather, the shared space instantiated by mirror neurons blends the interacting individuals within a shared implicit semantic content.

The self–other identity preexists and further parallels the self–other dichotomy. As convincingly shown by developmental psychology, the "being like me" analogy relies heavily on action and imitation of action, but is not confined to the domain of action. It is a global dimension that encompasses all aspects defining a life form, from its distinctive body to its distinctive affect. This global dimension covers a broad range of implicit certitudes we entertain about other individuals.

In the following sections I discuss many different forms of interaction, all contributing to the composition of the global experiential dimension we share with others. I will try to recompose all these multidimensional articulations of the self–other relationships within an integrated neuroscientific framework by introducing a new conceptual tool: the shared manifold of intersubjectivity.

3.5 Self–Other Identity and Empathy

Self–other identity goes beyond the domain of action. It incorporates sensations, affect, and emotions. The affective dimension of interindividual relations attracted the early interest of philosophers because it was recognized as a distinctive feature of human beings. In the eighteenth century, Scottish moral philosophers identified our capacity to interpret the feeling of others in terms of "sympathy" (see A. Smith, 1759/1976). But it was

during the second half of the nineteenth century that these issues acquired a multidisciplinary character when they were tackled in parallel by philosophers and the scholars of a new discipline, psychology.

"Empathy" is a later English translation (see Titchener, 1909) of the German word "*Einfühlung*." It is commonly held that *Einfühlung* was originally introduced into the vocabulary of the psychology of aesthetic experience by Theodore Lipps (1903a) to denote the relationship between a work of art and the observer, who imaginatively projects herself into the contemplated object.

However, the origin of the term is actually older. As pointed out by Prigman (1995), Robert Vischer introduced the term in 1873 to account for our capacity to symbolize the inanimate objects of nature and art. Vischer was strongly influenced by the ideas of R. Lotze, who already in 1858 proposed a mechanism by means of which humans are capable of understanding inanimate objects and other species of animals by "placing ourselves into them" ("*sich mitlebend ... versetzen*").

Lipps (1903b), who wrote extensively on empathy, extended the concept of *Einfühlung* to the domain of intersubjectivity, which he characterized in terms of *inner imitation* of the perceived movements of others. When I am watching an acrobat walking on a suspended wire, Lipps (1903b) notes, "I feel myself so inside of him" ("*Ich Fühle mich so in ihm*"). We can see here a first suggested relation between imitation (though "inner" imitation, in Lipps's words) and the capacity to understand others by ascribing to them feelings, emotions, and thoughts.

Phenomenology has further developed the notion of *Einfühlung*. A crucial point in Husserl's thought is the relevance he attributes to intersubjectivity in the constitution of our cognitive world. Husserl's rejection of solipsism is clearly epitomized in his fifth *Cartesian Meditation* (1953/1977, English translation), and even more in the posthumously published *Ideen II* (1989, English translation), in which he emphasizes the role of others in making our world "objective." It is through a "shared experience" of the world, provided by the presence of other individuals, that objectivity can be constituted. Interestingly enough, according to Husserl, the bodies of self and others are the primary instruments of our capacity to share experiences with others. What makes the behavior of other agents intelligible is the fact that their body is experienced, not as material object (*Körper*), but as something alive (*Leib*), something analogous to our own acting body as we experience it.

From birth onward the *Lebenswelt*, the world inhabited by living things, constitutes the playground of our interactions. Empathy is deeply

grounded in the experience of our lived-in body, and it is this experience that enables us directly to recognize others, not as bodies endowed with a mind, but as *persons* like us. Persons are rational individuals. What we now discover is how a rationality assumption—we consider others to be persons like us, therefore rational beings—can be grounded in bodily experience. According to Husserl, there can be no perception without awareness of the acting body. It should be added that the awareness of our acting body cannot be detached from the mechanisms presiding over control of actions (see also Gallese, 2000a,b).

The relationship between action and intersubjective empathic relations becomes even more evident in the works of Edith Stein and Merleau-Ponty. In her book *On the Problem of Empathy* (1912/1964, English translation), Stein, a former pupil of Husserl, explains that the concept of empathy is not confined to a simple grasp of the other's feelings or emotions. Empathy has a more basic connotation. The other is experienced as another being like oneself through an appreciation of *similarity*. An important component of this similarity resides in the common experience of action. As Stein points out, if the size of my hand were given at a fixed scale, as something predetermined, it would become very hard to empathize with any other types of hand that did not match these predetermined physical specifications. However, we can easily recognize children's hands and monkeys' hands as such despite their different visual appearances. Furthermore, we can recognize hands as such even when all the visual details are not available, even despite shifts in our point of view, and even when no specification of visual shape is provided. Even if all we can see are moving light-dot displays of people's behavior, we are not only able to recognize a walking person, but also to discriminate whether it is ourselves or someone else we are watching (see Cutting & Kozlowski, 1977). Since in normal conditions we never look at ourselves when we are walking, this recognition process can be much better accounted for by a mechanism in which the observed moving stimuli activate the observer's motor schema for walking, than solely by means of a purely visual process. This seems to suggest that our grasping of the meaning of the world doesn't *exclusively* rely on its visual representation, but is strongly influenced by action-related sensorimotor processes.

Merleau-Ponty in the *Phenomenology of Perception* writes:

> The communication or comprehension of gestures come[s] about through the reciprocity of my intentions and the gestures of others, of my gestures and intentions discernible in the conduct of other people. It is as if the other person's intention inhabited my body and mine his. (1945, English translation 1962, p. 185)

Self and other relate to each other because they both represent opposite extensions of the same correlative and reversible system *self–other*. The observer and the observed are part of a dynamic system governed by reversibility rules.

The shared intersubjective space in which we live from birth continues long afterward to constitute a substantial part of our semantic space. When we observe other individuals acting, facing their full range of expressive power (the way they act, the emotions and feelings they display), a meaningful embodied link among individuals is automatically established.

The discovery of mirror neurons in adult individuals shows that the very same neural substrate is activated when some of these expressive acts are both executed and perceived. Thus, we have a subpersonally instantiated common space. It relies on the neural circuits involved in the control of actions.

The hypothesis I am putting forward here is that a similar mechanism could underpin our capacity to share feelings and emotions with others. My proposal is that sensations and emotions displayed by others can also be empathized with, and therefore implicitly understood, through a mirror matching mechanism.[1]

3.6 The Shared Manifold Hypothesis

Throughout this chapter I have argued that the establishment of a self–other identity is a driving force in the cognitive development of more articulated and sophisticated forms of intersubjective relations. I have also focused on the mechanism that enables this identity to be created. I suggest that the concept of empathy should be extended to accommodate and account for all the different aspects of expressive behavior that enable us to establish a meaningful link between others and ourselves. This enlarged notion of empathy opens up the possibility of unifying under the same account the multiple aspects and possible levels of description of intersubjective relations.

As we have seen, when we enter into relations with others, there is a multiplicity of states that we share with them. We share emotions, our body schema, somatic sensations such as pain, etc. A comprehensive account of the richness of content we share with others should rest upon a conceptual tool that can be applied to all of these different levels of de-

1. For discussions relevant to this section, see vol. 2, ch. 13 by Jesse Prinz and the comments by Huesmann, vol. 2, ch. 19.6, p. 386.

scription, while simultaneously providing their functional and subpersonal characterization.

I introduce the *shared manifold* of intersubjectivity as this conceptual tool (see Gallese, 2001, 2003b). I posit that it is by means of this shared manifold that we recognize other human beings as similar to us. It is just because of this shared manifold that intersubjective communication, social imitation, and mind reading become possible. The shared manifold can be operationalized at three different levels: a phenomenological level, a functional level, and a subpersonal level.

The *phenomenological level* is the level responsible for the sense of similarity, of being individuals within a larger social community of persons like us, which we experience any time we are confronted with other human beings. It could be defined also as the *empathic* level, provided that empathy is characterized in the enlarged way I advocate here. Actions, emotions, and sensations experienced by others become meaningful to us because we can share them with others.

The *functional level* can be characterized in terms of *"as if"* modes of interaction that enable models of a self–other identity to be created. The same functional logic is at work during control of one's own actions and in understanding others' actions. Both are models of interaction that map their referents onto the same functional nodes and share a relational character. At the functional level of description of the shared manifold, its relational character produces the self–other identity by enabling the system to detect coherence, regularity, and predictability independently from their source.

The *subpersonal level* is characterized by the activity of a series of mirror matching neural circuits. The activity of these neural circuits is in turn tightly coupled with multilevel changes within body states. We have seen that mirror neurons instantiate a multimodal intentional shared space. My hypothesis is that analogous neural networks might be at work generating multimodal emotional and sensitive shared spaces—the shared spaces that allow us to appreciate, experience, and implicitly understand the emotions and the sensations we assume that others experience (see Goldman & Gallese, 2000; Gallese, 2001, 2003b). No systematic attempt has been produced so far to validate or falsify this hypothesis experimentally. Yet there are clues that my hypothesis might be not so ill founded.

Preliminary evidence suggests that in humans a mirror matching mechanism is at work in pain-related neurons. Hutchison et al. (1999) studied pain-related neurons in the human cingulate cortex. Cingulotomy procedures for the treatment of psychiatric disease provided an opportunity to

examine prior to excision whether neurons in the anterior cingulate cortex of locally anesthetized but awake patients responded to painful stimuli. It was noticed that a neuron that responded to noxious mechanical stimulation applied to the patient's hand also responded when the patient watched pinpricks being applied to the examiner's fingers. Both applied and observed painful stimuli elicited the same response in the same neuron.

Calder et al. (2000) showed that a stroke patient who suffered damage to the insula and the putamen was selectively impaired in detecting disgust in many different modalities, such as facial signals, nonverbal emotional sounds, and emotional prosody. The same patient was also selectively impaired in subjectively experiencing disgust and therefore in reacting appropriately to it. Once the capacity to experience and express a given emotion is lost, the same emotion cannot be easily represented and detected in others.

Emotions constitute one of the earliest ways to acquire knowledge about the situation of the living organism and to comprehend it in the light of its relations with others. This points to a strong interaction between emotion and action. We dislike things that we seldom touch, look at, or smell. We do not "translate" these things into motor schemas suitable for interacting with them (most likely "tagged" with positive emotions), but rather into aversive motor schemas (most likely "tagged" with negative emotional connotations). The coordinated activity of sensorimotor and affective neural systems results in the simplification and automatization of the behavioral responses that living organisms need to produce in order to survive.

The strict coupling between affect and sensorimotor integration is demonstrated in a study by Adolphs et al. (2000) in which these authors reviewed more than a hundred brain-damaged patients. Among other results, this study shows that patients who have suffered damage to sensorimotor cortices score worse than others when asked to rate or name facial emotions displayed by human faces.

Iacoboni and co-workers (Carr et al., 2001; see also Iacoboni, vol. 1, ch. 2) in a recent functional magnetic resonance imaging (fMRI) study on healthy participants showed that both observation and imitation of facial emotions activate the same restricted group of brain structures that includes the premotor cortex, the insula, and the amygdala. It is possible to speculate that such a double activation pattern during observation and imitation of emotions could be due to the activity of a neural mirror matching mechanism.

My hypothesis also predicts the existence of somatosensory mirror neurons that give us the capacity, when observing other bodies, to map

different body locations onto equivalent locations on our own body. New experiments on both monkeys and humans to test this hypothesis are just getting started in our laboratory.

It should be added that the shared manifold of intersubjectivity does not require that we experience others the same way we experience ourselves. Rather, the shared manifold enables and bootstraps mutual intelligibility. Self–other identity is not all there is to intersubjectivity. As pointed out by Husserl (1973), if this were the case, others could not be experienced as others (see also D. Zahavi, 2001). On the contrary, the *alterity* of the other grounds the objective character of reality. The quality and content of our own self-experience of the external world are constrained by the presence of other subjects who are intelligible while preserving their character as other. This alterity, as we have seen, is present also at the subpersonal level instantiated by the different neural networks coming into play when *I* act versus when *others* act.

3.7 Conclusions

There is preliminary evidence that the same neural structures that are active during sensations and emotions are also active when the same sensations and emotions are detected in others. It appears therefore that a whole range of different mirror matching mechanisms may be present in our brain. This mechanism, originally discovered and described in the domain of actions, is most likely a basic organizational feature of our brain.

One of the mechanisms enabling emotional feelings to emerge is the activation of neural "*as if* body loops" (Damasio, 1999). These automatic, implicit, and nonreflexive simulation mechanisms, bypassing the body proper through the internal activation of sensory body maps, create a representation of emotion-driven, body-related changes. It is likely that the activation of these "*as if* body loops" can not only be internally driven but can also be triggered by observation of other individuals (see Adolphs, 1999; Goldman & Gallese, 2000; Gallese, 2001).

The discovery of mirror neurons in the premotor cortex of monkeys and humans has unveiled a neural matching mechanism that, in the light of more recent findings, appears to be present also in a variety of nonmotor-related human brain structures. Much of what we ascribe to the mind of others when witnessing their behavior depends on the "resonance mechanisms" (see Rizzolatti, vol. 1, ch. 1) that their behavior triggers in us. The detection of intentions that we ascribe to observed agents and that we assume to underpin their behavior is constrained by the necessity for an

intersubjective link to be established. Early imitation is but one example of the intersubjective link in action. The shared manifold I have described here is a good candidate for determining and shaping this intersubjective link.[2]

Acknowledgments

Shorter and preliminary versions of this chapter were presented at the Munich Encounters in Cognition and Action at the Max Planck Institute in Munich, Germany, in December 2000; at the thirty-first annual meeting of the Jean Piaget Society, held in Berkeley, California, in May 2001; at the second meeting of the McDonnell Project in Philosophy and the Neurosciences, held in Tofino, Canada, in June 2001; and at the Royaumont conference on imitation, from which this book originated. I wish to thank all audiences for the feedback I have received from them. This work was supported by the Eurocores Program of the European Science Foundation and by Ministero dell 'universita' e della Ricerca scientifica e tecnologica.

2. See the comments on this chapter by Jones, in vol. 1, ch. 8.4, p. 205; see also ch. 7 by Hurley in vol. 1 and ch. 3 by Gordon in vol. 2. ED.