

# Neurosciences and Philosophy of Mind: A Reductive Interpretation of the “Mirror Neurons System” (MNS)

Nicola Simonetti\*

High School Teacher, PhD Student in Cognitive Science, University of Siena

\*Corresponding author: [simonettin@libero.it](mailto:simonettin@libero.it)

Received February 24, 2014; Revised March 11, 2014; Accepted May 14, 2014

**Abstract** The first group of reflections I want to develop is about the way in which the mind/body problem is placed within the sciences that, in virtue of a well-established convention, are unified under the name of “neurosciences”. I turn now to a second set of considerations that enter into the merits of mind/body problem. In order to this problem, like many other scholars, I support a thesis to which I give the name of “ontological monism”. From the ontological point of view, mental activity must necessarily (almost under an irrefutable postulate) be thought as a product of the work of the brain. The two neurobiological and mental conceptual systems may come into relation only if we place some ad hoc assumptions which work as a bridge between the two systems, allowing to relate psychological knowledge with neurobiological one. Another group of preliminary reflections concerns the distinction between possession and use of mental functions. In relation to higher mental functions (perception, attention, memory, language, thought, aimed behavior, etc.) I think very useful, both for theoretical investigation and for clinical distinction, between possession and use of the function. One of the richest and most exciting books in reporting experiments and implications arising from this important discovery is Rizzolatti and Sinigaglia (2006). The central argument around which the seven chapters of the book are articulated is that «the brain that acts is also and above all a brain that understands» (Ib., p.3). MNs, or “cerebral cells of empathy”, as they are sometimes called, are in these last years the focus of an important debate between many scientists, namely V.Gallese and A.Caramazza. Pascolo, anticipating in a study of 2008 the criticism of the research group of Caramazza about the existence or, at least, the cognitive role played by mirror neurons, criticized in a radical way the theories of Rizzolatti and his colleagues at the University of Parma about the existence and the role of mirror neurons. P. Jacob underlies the fact that MNs are not able to explain all the cognitive functions which are necessary to make possible the understanding of intentions in other people and the complex phenomenon of empathy. L. Boella, Professor of Moral Philosophy at the University of Milan, states that the discovery of mirror neurons has certainly contributed to the actual and popular success of neurosciences. From Boella’s point of view this success has been favoured by the special evidence and simplicity of the result of this discovery. By this way MNs discovery has spread through many disciplines, being interpreted beyond the specific context of the experimental research. MN activity, on my view, is more closely related to understanding goal-directed behavior than intentional actions. Now I’m going to show the theoretical concordance of this discovery with a reductive theory of mind, and particularly with the supervenience reductive theory of mind claimed by Kim (1993, 1996, 1998, 2005).

**Keywords:** *mirror neurons, neurosciences, philosophy of mind, supervenience, reductionism, mental causation, Rizzolatti, Gallese, Jacob, Pascolo, Caramazza, Boella, Kim*

**Cite This Article:** Nicola Simonetti, “Neurosciences and Philosophy of Mind: A Reductive Interpretation of the “Mirror Neurons System” (MNS).” *Research in Psychology and Behavioral Sciences*, vol. 2, no. 2 (2014): 24-42. doi: 10.12691/rpbs-2-2-1.

## 1. Introduction: Philosophical Preliminaries on the Mind-body Problem for Neurosciences and Cognitive Sciences

### 1.1. The Multi-layered Building of Neurosciences

The first group of reflections I want to develop is about the way in which the mind/body problem is placed within the sciences that, in virtue of a well-established convention, are unified under the name of “neurosciences”. The word

“neurosciences” refers to a broad spectrum of sciences which constitute a multi-layered building, with its foundations, with low and high levels. I will use this metaphor of the building to express my thoughts. Starting from the lowest level, we find, in ascending order (from the ground to the upper floors), sciences such as genetics, embryology, molecular biology, microscopic and macroscopic anatomy of the nervous system, neurophysiology, biochemistry of the brain processes, etc.

It’s in these disciplines that we have the evolving neuroscientific knowledge and discoveries, including, most recently, the very debated discovery of the so-called

“mirror neurons”. The knowledge of the brain and nervous system began to develop during the nineteenth century, and then in the last century and particularly in recent decades has progressed by leaps and bounds. The knowledge and discoveries in neuroscience I refer to are characterized by two important factors.

1) The use of highly sophisticated technologies made possible by the research based on strong science, like physics and chemistry. I allude to medical instruments as Transcranial Magnetic Stimulation (TMS), functional Magnetic Resonance Imaging (fMRI), Electro Encephalography (EEG), etc.

2) The systematic use of rigorous testing procedures – those ones of natural sciences - which usually exclude any possible ambiguity.

From the point of view of the nature and reliability of the neuroscience knowledge in these key disciplines mind/body problem do not pose any particular epistemological or methodological problems.

At this level of research it is given the assumption that mental activity depends on the work of the brain, and this assumption does not influence the investigation around the formation, morphology and functioning of the brain itself.

Let me explain: the problem of the nature of mental activity does not enter the field of inquiry in the study of genetics, anatomy or physiology of the brain. From their point of view, it is a problem that affects the speculative philosophical and not the scientific research, or at least marginally. Things start to get complicated if we continue to climb up the neurosciences building till the higher and top floors. Here we meet disciplines closely related to earlier ones, genetics, anatomy and physiology. They are based on them and yet are very different from the epistemological point of view, namely in terms of how they acquired the knowledge.

At the top of the building we meet: the neurobiology of cognition and emotion, neuropsychology, cognitive psychology, artificial intelligence, philosophy of mind, linguistics, etc. Scientists and scholars engaged in this research at the top of the neurosciences building do not deal with the important “small things” of our brain, but with the high functions and phenomena of our brain. Making use of the words of A. Lurija<sup>1</sup>, they deal with higher cortical functions: perception, attention, memory, recognition, thought, consciousness, emotions, the ability to plan and implement aimed behaviors. They would also deal with the realm of moral and artistic creativity.

Mind-body problem clearly emerges for these disciplines, not only as a speculative problem, but as a

practical issue that guides the theoretical investigation and clinical practice. I try to explain it briefly. In the theoretical research, as in the clinical one, neurosciences of the upper floors are interested in topics such as memory, thinking, emotions etc. It is now clear that in order to scientifically study these things you should know what kind of things they are. For example, to study the brain bases of memory, you need a reliable scientific concept of memory.

In general, when we study neurosciences from the point of view of higher mental functions, mind/body problem acquires a clear theoretical and practical importance for the simple fact that higher mental functions do not lend themselves to strictly scientific inquiry - as it happens, on the contrary, for disciplines such as genetics, embryology, anatomy, and so on. The question “why the higher functions are not likely to be treated scientifically?” opens a conversation without end. I limit myself to a banal observation: the mental functions are objects that I would define “multi-purpose”. In fact, they lend themselves to be treated and conceptualized in different ways, through different linguistic and conceptual systems.

There is not a universal criterion to describe mental functions. Various theories of the mind and its functions are sometimes different or even seemingly incompatible because maybe focus on different aspects or speak different languages. We are probably faced with an apparent paradox for scientific inquiry.

## 1.2. The Ontological Monism

I turn now to a second set of considerations that enter into the merits of mind/body problem. In order to this problem, like many other scholars, I support a thesis to which I give the name of “ontological monism”. We may explain it in the following way. From the ontological point of view, namely from the perspective of how things are in themselves, regardless of how we know, the thesis of ontological monism claims that mental activities and behaviors are products of the organism, and in particular of the central nervous system (CNS). The idea of a mental activity which takes place in the absence of a corresponding brain activity is unthinkable, is a “material nonsense”, as Husserl<sup>2</sup> would say. The physician and philosopher Pierre Cabanis<sup>3</sup>, a member of the movement of the late Enlightenment ideologues, liked to say that as

<sup>2</sup>E.G.A. Husserl (1859-1938) was an Austrian-German philosopher and mathematician, founder of phenomenology, and a member of the School of Brentano. The current philosophy of phenomenology has influenced much of European culture of the twentieth century and beyond.

In addition to M. Scheler's philosopher phenomenology had a profound influence on the existentialism and on M. Heidegger's philosophy. Finally, his thinking has influenced indirectly on a certain part of the Cognitive Science and Philosophy of Mind today (for example, according to H. Dreyfus, Husserl is considered the “father of contemporary research in cognitive psychology and artificial intelligence” as he endorses in 1982 *Husserl, Intentionality and Cognitive Science*, Bradford/M.I.T. Press, 1982.

<sup>3</sup>P. J. G. Cabanis (1757-1808) was a French physician, physiologist and philosopher. According to Cabanis, training is conducted by the ideas of our organic sensibility, that also direct the activities of our organs, and thus the totality of all living things. From the observation of pathological conditions, or the effect of narcotics and associated psychological states, he presents our thoughts as physiological results of a perception by an appropriate body, the brain.

In this way, Cabanis connected the instinct to the material structure of every living being, just as every organ has an its own predisposition to do this or that specific task in the body.

<sup>1</sup>A. R. Lurija (1902-1977) was a physician, psychologist and sociologist in the Soviet Union. He was disciple and collaborator of Lev Vygotskij, who was one of founders of the historical-cultural school and neuropsychology. The architecture of the brain functions, that Lurija discovered, consists of three major systems or functional units:

1) the first regulates the sleep-wake cycle and modulates cortical activity on the attention, the selection of information and perception of emotions; 2) the second has the primary function in perception, analysis and memory, and it is concerned with temporal, parietal and occipital cortices; 3) the third provides for the regulation, modulation and control of voluntary actions, affecting the frontal and prefrontal cortical and motor areas, cerebellum and deep nuclei.

These functional units are not, according to Lurija and Vygotskij too, genetically determined, but ontogenetically determined, through the pressure of the historical and cultural factors, and therefore they take on different characteristics depending on the periods and contexts of history and human society.

the stomach secretes gastric juice, so the brain secretes thought. This statement contains, in a provocative version, the essence of ontological monism that we are arguing here.

If there is no brain function, there can be no perceptions, thoughts, emotions and other mental activities. Before concluding with the ontological point of view, I'll try to deepen its philosophical meaning. The idea of a mental activity which unfolds on its own, in the absence of a brain function which produces it, cannot be the object of a scientific thought. Similarly, and for the same reasons, we cannot think that a chair is able to speak or a toy to feel emotions. These things can happen only in fairy tales. They can be produced by the imagination and therefore as images, sometimes very seductive, but nothing more.

Theology and metaphysics take the freedom to believe in the existence of an immaterial soul, capable of thinking and feeling even after the separation of the body. The metaphysics realm of images and symbols is so compelling as far from scientific thought. In other words, the theory of an ontological dualism, in the manner Descartes<sup>4</sup> or, in our times, John Eccles<sup>5</sup> and other few neuroscientists and philosophers, is not a product of good way of reasoning, while blanketing a scientific language, but it is generated by a science fiction capable of producing certainly seductive, but not rational images and concepts.

### 1.3. The Epistemological Dualism

The third point of reflections concerns the epistemological dualism. It's a thesis in which I firmly believe and that in different versions is supported by many scholars. From the ontological point of view, as I argued before, mental activity must necessarily (almost under an irrefutable postulate) be thought as a product of the work of the brain. Which is to say that it is unthinkable that a cognitive or phenomenal mental process or quale (reasoning, perception, intention, emotion, etc.) occurs in the absence of and adequate brain activity which generates it.

But if we change the point of view, departing for a moment from the objective point of view to the subjective one, things abruptly change in a radical way, because we are "dualist beings", in spite of the irrefutable ontological monism which identifies mental states with brain states. The crucial point, in my point of view, is that in the subjective thought and language neurobiology cannot work because every knowledge and representation of our inner mental states take place in the domain of psychology.

So, if we take into account both the way in which we know and act on things, and the way in which they are

physically realized and scientifically described, we are obliged, from my point of view, to endorse a position that I (and many scholars) call "epistemological dualism" or "duality of knowledge". I try to explain briefly this crucial concept. The conceptual and linguistic framework and methods by which we know and describe, through neurosciences and cognitive sciences, the structure and functioning of the brain and NS are completely different from the conceptual and linguistic system that we use to understand and describe the mental activity and behavior. They are completely different in the sense that the two conceptual and linguistic systems are not translatable one into the other. This untranslatability makes arise hard problems in the assumption *prima facie* of a full reduction of the mental domain to the physical domain.

The untranslatability of the two systems can be described in the following way: the concepts we use to describe a brain process are not suitable (Wittgenstein<sup>6</sup> would say that "they spin freely") to describe the corresponding mental process. Vice versa, the concepts we use to describe a mental process spin freely when applied to the underlying brain process. According to Wittgenstein's language games theory I cannot say in a precise way, for instance, that Broca's area (the area of the brain, placed in the base of the left frontal lobe, which physically realizes the ability to speak) speaks because this ability and its features belongs to another conceptual and language grammar.

Another example. It's universally accepted biological knowledge that the intellectual activities heavily depend on the functioning of the frontal lobes. But it would be meaningless to assert that the frontal lobes think. It's clear that while it's correct to say that frontal lobes make possible pur thinking, it's at all incorrect to say that frontal lobes think. For it is very important, as it may be seen in these two simple examples, the distinction between possession and use of a mental function.

In the Broca's area physical and biochemical processes take place and they make it possible to speak, while it would be a nonsense to say that Broca's area speaks. Coming back to Wittgenstein arguing, it is not logically possible to apply the mentalistic concept of "speaking" to the neurobiological concept of "Broca's area". The same applies to the thinking and the frontal lobes, and so on. But if it isn't Broca's area to speak, who is speaking? This is a typical way of arguing that Descartes and a cartesian thinker would endorse. In this case and in every case like this one, the correct answer is that the predicate of speaking or thinking, etc., precisely refer and belong to a human being or some other animal. This is the classic philosophical problem of the identity (who am I?) and, leaving aside the many theories about it during all the history of philosophy, we may see that it is certainly a

<sup>4</sup> R. Descartes (1596-1650) was a French philosopher and mathematician. Considered the founder of modern philosophy and the father of analytic geometry. In particular, Descartes argued an interactionist dualism between mind and body, by supporting that the mind ("res cogitans") interacts with the body ("res extensa") through the pineal gland.

<sup>5</sup> Sir J. C. Eccles (1903-1997) was an Australian neurophysiologist. He was the author of fundamental discoveries on-the physiology of neurons (nerve cells) and in particular the biochemical mechanism of nerve impulses. In clearing of this ion mechanism, leading to excitation and inhibition underlying physiology of the nervous system, the contribute of the British A. L. Hodgkin and A. F. Huxley, with whom they shared the Nobel Prize in Medicine and Physiology in 1963, was also important. Together with others few neuroscientists and philosophers he endorses a mind/body dualism.

<sup>6</sup> L. J. Wittgenstein (1889 – 1951) was an Austrian-born philosopher who held the professorship in philosophy at the University of Cambridge. Wittgenstein inspired two of the century's principal philosophical movements, logical positivism and ordinary language philosophy, though in his lifetime he published just one book review, one article, a children's dictionary, and the 75-pages *Tractatus Logico-Philosophicus* (1921). His *Philosophical Investigations*, published posthumously (1953) is considered one of the most important books of 20th-century philosophy. In this last book he provided a detailed account of the many possible uses of ordinary language, considering language as a series of interchangeable language-games in which the meaning of words is derived from their public use.

mentalist concept, perhaps the most mentalistic one, from which every other mentalistic concept takes its origin.

By this way we come to this actual conclusion, which many scholars share: although the mind is a product of the brain, the conceptual system that allows us to know and speak of the brain is immeasurably different from the conceptual system that we use to study and learn about the mental activities. This leads to two distinct domains of knowledge: the neurobiological domain and the psychological/behavioral domain. There is not, for the over logical and linguistic reasons, the possibility to pass directly from one to the other unless we argue, as the American philosophers Paul and Patricia Churchland<sup>7</sup> endorse, stating that in a near future conceptual and linguistic vocabularies of philosophy and neurosciences will mix together to create a new science of the mind, the neurophilosophy.

This relation between the neurobiological domain and the psychological/behavioral domain, or between neurosciences and philosophy, brings us to the fourth set of observations.

#### 1.4. The *ad Hoc* Theories or Bridge Theories

The two neurobiological and mental conceptual systems may come into relation only if we place some *ad hoc* assumptions which work as a bridge between the two systems, allowing to relate psychological knowledge with neurobiological one. The use of *ad hoc* assumptions is perfectly legitimate as long as we are aware of using them. I'll make an example for the concept of consciousness. Wittgenstein, in his *Philosophical Investigations* (1953), as I wrote before, said that the meaning of a word is the use we make of it. Therefore, he argued that a word has different meanings depending on the context of communication (language game) in which the word occurs. Now we have to wonder: how many different language games there are for the word and concept of consciousness? There are certainly some language games for it: we may speak of consciousness to refer to moral responsibility and free will, or we may speak of consciousness to refer, in a minimal way, to a simple biological state of awareness, etc.

But, despite the different uses in which "consciousness" may occur, the different meanings of consciousness are based on one single and basic concept, which is the ontological concept, the biological one, without which we couldn't assign any other concept. Nevertheless, it's also clear that the everyday concept of "consciousness", relating to moral responsibility and free will, etc., is essential for our language communication. And in everyday conversation misunderstandings or different opinions about its meaning rarely arise. The places for the disputes of meaning are almost always the discussions between philosophers and scientists. Therefore, it's clear that the concept of consciousness cannot be totally

sacrificed because we would deprive our language of an essential piece.

Vice versa, in the scientific and clinical domains some uses or aspects of the concept of consciousness can be sacrificed. This leads us to the *ad hoc* theories that are being introduced to establish a bridge between the mental and the brain. Let's examine one example to understand better this important account. In neurology, the concept of consciousness is characterized in terms of awareness. This characterization allows a strict translation of the mentalistic concept of consciousness in a rigorous neurological device. The neurologist defines consciousness in a negative from the detailed description of situations characterized by a decrease or absence of consciousness. To understand better what I'm going to say, consider two situations.

First situation. A person who is asleep and we add, not to complicate things too much, that it is a non-REM sleep. This person is unconscious in two main ways: i) within certain limits, he does not respond to external stimuli; ii) he has a characteristic ElectroEncephaloGraphic trace (EEG).

Second situation: a person in a deep coma. In this situation the absence of consciousness is defined by specific clinical and instrumental criteria. On a clinical level the main features are: a prolonged state of unconsciousness, like sleep, but marked by a complete non-responsiveness to external stimuli, even the most energetic, with a complete abolition of sensibility and movement, and serious vegetative disorders. On the instrumental level we have an EEG which shows a profound impairment of brain electrical activity, along with other important physiologic changes instrumentally tested. Therefore it is correct to say that in neurology a person has got a consciousness if he is not in one of these states or in other similar ones, such as fainting or syncope.

Here we have a precise and rigorous example of translation from the mental to the biological. But it's clear that the concept of consciousness used by the neurologist is a concept aimed solely and exclusively to the needs of neurology. That is precisely an *ad hoc* concept. The neurologist adopt a theory *ad hoc*, or a bridge theory of consciousness. Consciousness is translated by neurologist in a well-defined NS and body state. He has no interest for all other aspects or conceptual and linguistic uses of "consciousness" and he may ignore or sacrifice them. For example, he is not interested in the fact that we often use in our everyday language the word "moral consciousness" to refer to responsibility, or that sometimes we use the word "consciousness" to say that someone is very conscientious or not conscientious at all.

It is clear that the moral concept of "conscientiousness" used in these language games has nothing to do with the biological concept of "consciousness" in which neurology is basically interested. This is a clear example of a rigorous conceptualization of the *ad hoc* concept of consciousness: some aspects (those that are useful for the neurologist) are themed, the others, which are not considered interesting for the neurologist's purposes, are discarded. In other cases, *ad hoc* theories may be used in a less rigorous, absolutist way. I'll take an example from the field of psychopathology.

Consider a delirium. A delirium is a false belief that is supported even in front of the most obvious evidence of its falsity. Delirium is a serious mental symptom just

<sup>7</sup> Patricia S. Churchland has focused his philosophical investigation on the interface between neuroscience and philosophy. According to her, philosophers are increasingly realizing that to understand the mind one must understand the brain and in the near future neurosciences together with philosophy will create a new language with a neuroscientific base, thanks to the many discoveries. Churchland, Patricia Smith (2002). *Brain-Wise : Studies in Neurophilosophy*. The MIT Press. Churchland, Patricia Smith (1989) *Neurophilosophy : Toward a Unified Science of the Mind-Brain*. The MIT Press.

because it expresses a deep alteration of the subject's relationship with reality. It is possible, at least in part, to conceptualize the delirium to enable its translation into the language of neuroscience. Delirium is a mental phenomenon and as such can not be entirely described in an objective way. Nevertheless, through an appropriate *ad hoc* theory, it can be described as something objective. It is sufficient, in describing it, to take into account its legitimately objectivizable goals, rejecting everything appears refractory to objectivity. A delirium has an onset and its course. For example, it can occur at twenty or sixty. And these are objective data. Other objective data are about the duration of delirium, if it is a chronic, persistent delirium, or an episode delirium which appears and disappears in a short time. These are objective data too. But there are many others objective data about it: if delirium is confused or organized, if it's weird, if it is shiny, it is a delirium of persecution, or omnipotence, if it is responding to antipsychotic drugs (and which) or resistant to each drug. We can also take into consideration the informations of the patient's medical history and family: age, sex, social status, past illnesses and so on.

On the base of this *ad hoc* conceptualization, the delirium can be treated in the research such as a precise objective entity, which can be classified and investigated with statistical and epidemiological instruments, and it could become the subject of a research around the related neurobiology. This way of considering this mental pathological phenomenon has important advantages for theoretical research and for psychiatric practice, but there is a cost to be paid.

Translating delirium in an objective entity means to completely overlook the subjective side of the delirium experience, its being a personal life story, as something private and unrepeatable. The reification of the delirium involves the waiver of the psychological and existential search for meaning that it expresses. If on one hand the research can clearly go in the direction of the objectivity, on the other hand, it abandons the dimension of subjectivity and story of life.

In short, the *ad hoc* theories conceptualize mental activity with the aim of making it fit for empirical research and, where possible, to establish a bridge with the basic neuroscience research. This is perfectly legitimate and right, provided you know that you are sacrificing many aspects of the mental activity. The risk is that the *ad hoc* theory can become for the community of researchers the real and only truth.

## 1.5. Possession and Use of Mental Functions

Another group of preliminary reflections concerns the distinction between possession and use of mental functions. In relation to higher mental functions (perception, attention, memory, language, thought, aimed behavior, etc.) I think very useful, both for theoretical investigation and for clinical distinction, between possession and use of the function. Take for example the function of the linguistic expression. The possession of this function depends on the level of the brain, the integrity of certain cortical areas well known in neurobiology, including Broca's area.

If these structures suffer a damage, the expressive function is altered. The person loses all or part of the ability to speak. Here you have the possession of the

function, which may exist or be questioned in the case of a brain damage. Possession is entirely in the domains of neurobiology and neuropsychology. But in addition to the possession of a function there is the use we make of it. The use of a function has a more qualitative feature. For example, the full possession of the expressive function of language is consistent with the fact that the subject can make a good or a bad use of the words. In fact we can speak saying something inappropriate or stupid, helpful or harmful things, things you understand or that are unintelligible.

This applies to the language, but also to all the other higher functions: memory, reasoning, learning, and so on. There's no doubt that also the quality of the use of a function depends (according to the theory of ontological monism) on the brain and NS, but the at the present state of neurobiology there is no satisfactory answer to the question of the use of a function. I mean that, for example, the brain of a Nobel prize for literature and the brain of a healthy illiterate man usually do not show any difference in relation to the present acquisition and tools of neurobiology.

And if the illiterate man is young, and the Nobel Prize is old, the brain of the latter will certainly be less healthy. This reflection is useful to make clear the fact that the use of function calls into question a wide variety of variables that are largely foreign to the issue of possession, that is to the integrity of the organ. It's clear that the problem of the use is correspondingly more difficult and elusive than the problem of possession.

## 2. The Mirror Neurons System (MNS)

### 2.1. A Short History of the Discovery

I'm going to start considering mirror neurons discovery with a short selection of many recent quotes from well-known neuroscientists and philosophers who are writing and debating very much on this discovery, its interpretation and its consequences for our theories of mind and cognitive functions.

The mirror neurons are for psychology what DNA was for biology.<sup>8</sup>

The discovery of motor resonance mechanism of mirror neurons has shown that the motor system, far from being a mere muscles controller and a simple executor of commands coded elsewhere, is able to perform cognitive functions that for a long time have been erroneously considered prerogative of psychological processes and neural mechanisms of a purely associative kind ... Surely mirror neurons bother who look at neuroscience as a simple method of tracking and validation of mental mechanisms deemed valid a priori.<sup>9</sup>

It is important to keep several questions distinct. When do we have simulation? When do we have mirror neurons? When do we have social-intentional relationships? I do not

<sup>8</sup> (Vilayanur S. Ramachandran, professor of Neuroscience and Psychology at the University of California, San Diego, director of the "Center for Brain and Cognition" and is an adjunct professor of biology at Salk Institute).

<sup>9</sup> (Vittorio Gallese, Medical Neurologist, is Professor of Physiology at the University of Parma. He was part, with Giacomo Rizzolatti, Leonardo Fogassi, of the research group known throughout the world to have discovered "mirror neurons", along with Luciano Fadiga, now at the University of Ferrara).

argue that conceptualization and imputation are necessary for the existence or activation of mirror neurons (MNs), only that they (or something similar) are necessary for social-intentional relationships. If MNs themselves do not guarantee such elements, then they don't, all by themselves, guarantee social-intentional relationships.<sup>10</sup>

Human ability to represent psychological states (beliefs, intentions, desires, emotions) and to attribute them to others (so-called mindreading) goes beyond the mechanism of mirror neurons. Consequently, also the idea that autism stems from lack of mirror neurons is wrong.<sup>11</sup>

Right now it has not been demonstrated that mirror neurons play a very functional role in understanding the action. And, even if they do, how they do it.<sup>12</sup>

In the '80s and '90s the group of researchers at the University of Parma headed by Giacomo Rizzolatti and composed by Luciano Fadiga, Leonardo Fogassi, Vittorio Gallese and Giuseppe di Pellegrino was devoted to the study of the pre-motor cortex. They had placed electrodes in the inferior frontal cortex of a macaque monkey to study the specialized neurons in the control of hand movements, such as collecting or handling objects. During each experiment it was recorded the behaviour of individual neurons in the monkey brain while it allowed the macaque to access bits of food, in order to measure the neural response to specific movements. Like many other important discoveries, that of mirror neurons was due to chance.

The anecdotal reports that while one investigator took a banana in a fruit basket prepared for the experiments, some neurons in the so called F5 brain area and around (in which there are visuo-motor neurons) in the monkey, which was watching the scene, had reacted. How could it happen if the monkey had not moved? How could it happen if until then we thought that these neurons are activated only for motor function? At first, investigators thought it was a defect in the measures or a failure in the instrumentation, but everything turned out okay and the reactions were repeated as soon as it repeated the action of grasping.

Since then, this work has been published, with the update on the discovery of mirror neurons located in both the inferior frontal parietal regions of the brain and confirmed. In 1995, Luciano Fadiga, Leonardo Fogassi, Giovanni Pavesi and Giacomo Rizzolatti demonstrated for the first time the existence in humans of a system similar to that found in monkeys. Using Transcranial Magnetic Stimulation (TMS), they found that the human motor cortex is facilitated by the observation of human actions and movements of others.

More recently, other evidences obtained by functional Magnetic Resonance Imaging (fMRI), TMS, ElectroEncephaloGraphy (EEG) and behavioural tests have confirmed that similar systems existing in the human brain are highly developed. They have been precisely identified the regions which respond to the action/observation. Given the genetic similarity between

primates (including humans), it is not surprising that these brain regions are closely similar in them.

During the last years researchers thought and created more and more sophisticated experiments on monkeys. Particularly, I want to remind and mention a recent ingenious experiment. The usual criticsians had argued that the alleged "mirror neuron" is only sensitive to a movement of grabbing or handling or tighten the fingers, not to the intention of the action with a clear and precise purpose. To show that the activation of the neuron is also sensitive to the purpose, researchers used the typical pliers for French gourmet escargotes. These pliers open when you tighten the grip and viceversa close when you release the grip. Training a monkey to grab a nut with these pliers, you may see the activation of the mirror neuron which fires specifically when there is an action of grabbing of food to eat it. The fingers are relaxed rather than contracted, but the intentional action, on an abstract level, is the same and therefore the mirror neuron fires. This interesting and important experiment should clearly show the fact that MNs are not only sensitive to a movement of grabbing or handling or tighten the fingers, but also to the intention of the action with a clear and precise purpose.

Finally, at the beginning of April 2010, the team of neuroscientists headed by Roy Mukamel and Marco Iacoboni, University of California at Los Angeles (UCLA), gave the important new that the problem of proving in a direct way the existence of such a mirror neurons system in humans (because it is considered unethical to implant electrodes in the brain of people for these research purposes), has been overcome, thanks to twenty-one patients treated for epilepsy which was resistant to medication (reference: Mukamel et al., Single-Neuron Responses in Humans during Execution and Observation, *Current Biology*, 2010).

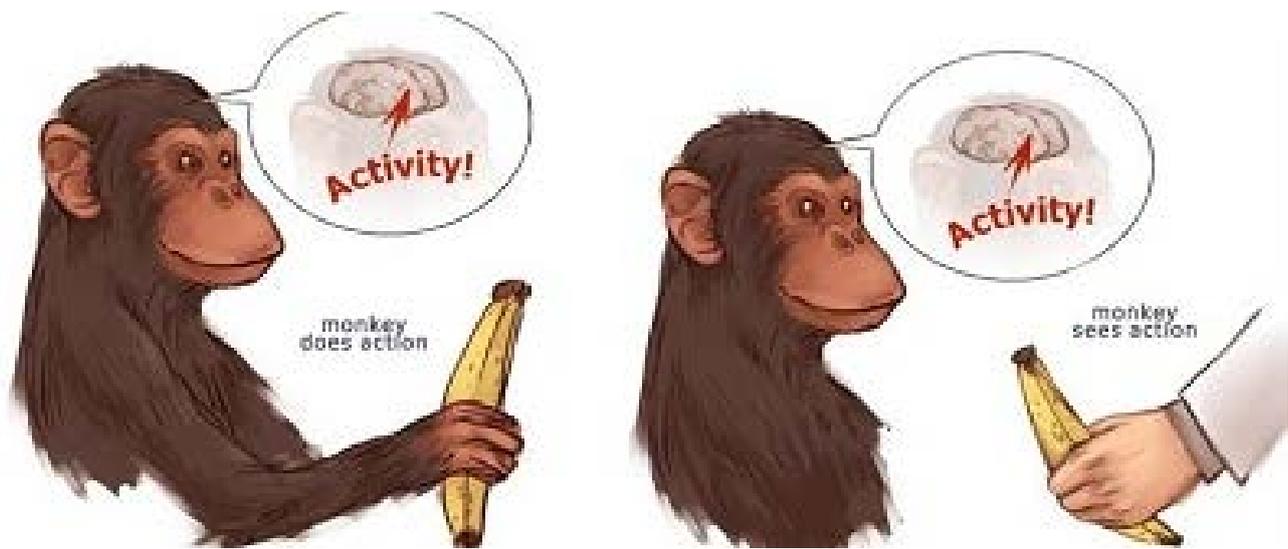
Some electrodes have been planted in their brain with the aim to identify the epileptogenic foci for a neurosurgical intervention, assuming that the electrodes placement was based solely on clinical criteria. During their hospitalization, the researchers told them to perform certain actions, such as grasping objects, or to observe facial expressions on a screen and then to repeat them. The authors say that among the 1177 neurons under observation a significant portion responds to both stimuli (action and observation of grasping and facial emotions) in the supplementary motor area of the medial frontal lobe (SMA) and in the medial temporal lobe, specifically in the hippocampal and the entorhial cortex, while in the amygdala and in the pre-SMA the number of cells would not reach significant levels of response.

According to the theory of MNS, mirror neurons (MNs) fired both for action and observation, and their excitement was directly recorded by the electrodes. These findings suggest, according to these researchers, the idea that in humans there are probably multiple systems with mirror neural mechanisms both for the integration and for the differentiation of perceptual and motor aspects for the actions performed by themselves and by others. Since these new areas of the cortex perform different functions (vision, movement, memory), Iacoboni thinks that this discovery suggests us the idea that mirror neurons provide a very rich and complex system of pre-logic reproduction and interpretation of actions and emotions of others.

<sup>10</sup> (Alvin Goldman, Board of Governors Professor of Philosophy and Cognitive Science at Rutgers, The State University of New Jersey).

<sup>11</sup> (Pierre Jacob, Philosopher of Mind and Cognitive Scientist, currently President of the European Society of Philosophy and Psychology and Director of the Jean Nicod Institute in Paris).

<sup>12</sup> (Alfonso Caramazza, director of the Laboratory of Cognitive Neuropsychology at Harvard University and director of the Center of the Mind-Brain of the University of Trento).



## 2.2. G. Rizzolatti and C. Sinigaglia's Interpretation of the MNS pp.14-19

One of the richest and most exciting books in reporting experiments and implications arising from this important discovery is Rizzolatti and Sinigaglia (2006). The central argument around which the seven chapters of the book are articulated is that «the brain that acts is also and above all a brain that understands» (Ib., p.3). The meaning and scope of this statement lie in the heart of the neural mechanism identified by the neurophysiologists at the University of Parma headed by G. Rizzolatti.

As I already said in the history of this discovery, in a series of studies conducted over the past two decades, these researchers discovered in the pre-motor cortex of monkeys and later also in the human one, by medical instruments of cerebral imaging, the existence of two groups of neurons which are both active during the implementation of actions related to objects: they are simple and familiar gestures like grabbing something with your hand or bringing food to your mouth.

The surprising thing is that these two groups of pre-motor neurons are also activated in the absence of any enforcement action during purely observational explicit tasks: the first group of neurons respond to the vision of the object to which the action could be directed, while those of the second group respond to the observation of another individual who performs the same action. Following the authors, we may take the example of the coffee cup: the pre-motor neurons are activated while you grasp the handle, but for some of them activation is triggered even by the simple observation of the cup resting on the table, for others also by the observation of our neighbour who grabs the cup to drink his coffee. Therefore we have in both cases bimodal neurons, that activate both for motor and perceptual processes.

Their activity may be described by the mechanism of "neural embodied simulation": during the observation of an object it activates a motor pattern appropriate to its characteristics (such as size and orientation in space) "as if" the viewer enters into interaction with it. In the same way, during the observation of an action performed by another individual, the neural system of the observer activates "as if" he were to perform the same action he observes. The neurons of the first group were called

"canonical neurons" because since the 30s it had suggested the involvement of the pre-motor areas in the processing of visual information about an object in the motor acts required to interact with it; those of the second group were called "mirror neurons" because they cause a mirror reaction in the neural system of the observer, in which it takes place a simulation of the observed action.

In the light of this mechanism of neural embodied simulation, it could be reinterpreted the role played by the motor system within the whole cognitive system, because the first was usually connected only with the planning and execution of actions. On the contrary, it seems that bimodal neurons found in pre-motor cortex are strongly implicated in high-level cognitive processes, particularly in the perceptual recognition of objects and actions, and in the understanding of their meaning. This new way of seeing and explaining the motor system, which comes to be also implied in perceptual recognition of objects and actions, and in the understanding of their meaning, undermines the rigid boundary between the perceptive cognitive processes and the motor ones. This rigid boundary between motor and cognitive processes has for years characterized the interpretation of the architecture of the brain. On the contrary, it seems that perception, understanding and action are grouped together into a unified mechanism, according to which «the brain that acts is also and above all a brain that understands» (Ib., p. 3).

Brain understanding regarding objects is related to their functional significance or "affordance". Canonical neurons allow an immediate understanding of the possible interactions that some objects have for a perceiving subject (in the case of the handle of a coffee cup, the possibility to be grasped). With regard to the actions the understanding is related to the purpose behind them. Instead mirror neurons enable an immediate understanding of the intentions of other individuals (e.g. the intention of a man to bring the cup to his mouth to drink the coffee), making possible a prediction of their future behaviour.

Many experiments were conducted on monkeys and humans to achieve these theories. Obviously the techniques used for monkeys and for humans are usually very different (except for the experiment of Mukamel and Iacoboni of April 2010): while in monkeys it's possible to make a record of a single neuron via intra-cortical

insertion of electrodes, in human subjects only non-invasive methods of brain imaging are used, such as Positron Emission Tomography (PET) or fMRI that allow us to visualize the activity of whole brain areas but not of individual nerve cells. This is, so far, the invincible limit of such experiments on humans.

Particularly in the fourth chapter of Rizzolatti and Sinigaglia (2006), entitled "Action and understanding", there are two experiments which are considered very important to define the role of mirror neurons in our understanding of the purpose underlying the actions. The former has revealed the existence of a mechanism not only in motor and visual mode, but also in auditory mode. Indeed, when the monkey is in darkness and listens to the noise produced by an action as breaking a nut, the same neuron fires when the animal breaks the nut, when it sees someone breaking a nut, and when it hears the sound of someone who breaks it. The interpretation of this experiment is that, whatever the mode, the same neuron fires to encode the "breaking-of-a-nut" that coincides with the purpose, that is the intention of the action.

The second experiment has allowed to discriminate between a gesture of grabbing aimed to bring the food to the mouth or put it in a container. This experiment goes in the same direction of that one on the typical pliers for French gourmet escargotes I talked about before. I talked about before During the execution of that precise action (grasping), MNs fired in different ways depending on the ultimate goal of action, that is if the intention was to bring food to the mouth or to move it into the container. In the same direction they seem to go some results obtained with humans by an experiment with fMRI. It was noted a particularly significant activation of the mirror system in experimental subjects during the observation of actions which were not "pure", but precisely included in the context, from which one could clearly infer the intention that was implied.

All these experiments would allow us to state that the mirror neuron system is able to encode not only the act, but also the intention with which it is made. According with the paradigm of the embodied cognition (endorsed by many philosophers and neurobiologists, namely A. Clark, A. Damasio, etc.), the intentions of others can be understood without any reflective conceptual or linguistic mediation. It would be nothing but a pragmatic understanding based solely on the motor knowledge on which it depends our own capacity to act.

Another very interesting chapter of the book is the sixth, entitled "Imitation and language". Two other important functions assigned to the mirror system are described, as basilar capacities which would make the verbal and non verbal language possible. They are: 1) an "imitative function" intended as the ability to replicate gestures already belonging to our own motor repertoire; 2) the "capacity of learning new motor patterns by imitation". It is a common function that would also outline a possible scenario for the origin of human language related to the evolution of the mirror system, which could be interpreted in this sense as an instrument of integration between gestures and sounds to foster a more precise understanding of social behaviours.

*Dulcis in fundo*, the last chapter of the book is dedicated to the sharing of emotions. The central thesis is that the recognition of the emotions of others is based on a

set of different neural circuits which share the mirror properties already seen in the case of action understanding. It was possible to study experimentally some primary emotions such as pain and disgust, and the results clearly show that observing in the other an expression of sorrow or disgust activates the same neural substrate underlying the perception in first person of the same kind of emotion, as if it were a sort of involuntary perceptual and motor imitation.

Further confirmation comes from clinical trials in patients suffering from neurological diseases. Once lost the ability to feel and express a given emotion, it becomes impossible to recognize even when expressed by others. As in the case of the actions, also for the emotions one can speak of an immediate, pre-logic understanding that do not require cognitive processes of the kind of conceptual inference or association. This immediate understanding of the emotions of others would be the necessary precondition for that empathic behaviour underlying much of our inter-individual relations. Moreover, as the authors rightly note, already Darwin himself (in *The Expression of Emotions in Man and Animals*, 1872) has emphasized the adaptive value of emotions and the evidence of the perceptual and emotional empathy in the animal kingdom.

Far from being confined to the functioning of certain nerve cells, the mirror properties would pervade the entire system of the brain: the same logic that allows us to pair execution and action understanding in a single neural mechanism, allows us also to describe the emotional sharing and perhaps the arising of the phenomenon of consciousness too. The neuro-psychologist Anna Berti has identified a similar mode of "neural coupling" for the execution of actions and the awareness of having (or not) performed them. This motor awareness, that allows us to be conscious of our actions would share the same neural substrate underlying the motor control of these actions.

### 3. The Debate Around Mirror Neurons (MNs) pp

#### 3.1. How to Interpret the New Results and Challenges of Neurosciences? V. Gallese Versus A. Caramazza (May 29, 2009)

"The discovery of motor resonance mechanism of mirror neurons has shown that the motor system, far from being a mere muscles controller and a simple executor of commands coded elsewhere, is able to perform cognitive functions that for a long time have been erroneously considered prerogative of psychological processes and neural mechanisms of a purely associative kind... Surely mirror neurons bother who look at neuroscience as a simple method of tracking and validation of mental mechanisms deemed valid a priori."

(Vittorio Gallese, Medical neurologist, is Professor of Physiology at the University of Parma)

"Till now we haven't had any important clinical demonstration that mirror neurons really have a functional role in the action understanding and, even if they do, in which way they can do it."

(Alfonso Caramazza, Director of the Laboratory of Cognitive Neuropsychology at Harvard University and

Director of the International Center for Mind/Brain Science of the University of Trento).

MNs, or “cerebral cells of empathy”, as they are sometimes called, are in these last years the focus of an important debate between many scientists, namely V. Gallese and A. Caramazza. As I told in the Chapter 2, MNs were discovered in the ‘90s by the group of researchers at the University of Parma headed by G. Rizzolatti and composed by L. Fadiga, L. Fogassi, V. Gallese and G. di Pellegrino during their experiments on pre-motor cortex of macaque monkeys brain, while they were collecting or handling objects.

Finally, at the beginning of April 2010, the team of neuroscientists headed by Roy Mukamel and Marco Iacoboni, University of California at Los Angeles, gave the important new that the problem of proving in a direct way the existence of such a mirror neurons system in humans has been overcome, thanks to twenty-one patients treated for epilepsy. Some electrodes have been planted in their brain for medical purposes. During their hospitalization, the researchers told them to perform certain actions, such as grasping objects, or to observe facial expressions. According to the theory of MNS, MNs fired both for action and observation of actions, and their excitement was directly recorded by the electrodes.

The authors say that among the 1177 neurons under observation a significant portion responds to both stimuli (action and observation of grasping and facial emotions) in the supplementary motor area of the medial frontal lobe (SMA) and in the medial temporal lobe, specifically in the hippocampus, parahippocampal and the entorhinal cortex, while in the amygdala and in the pre-SMA the number of cells would not reach significant levels of response. Since these new areas of the cortex perform different functions (vision, movement, memory), Iacoboni thinks that this discovery suggests us the idea that mirror neurons provide a very rich and complex system of pre-logic reproduction and interpretation of actions and emotions of others.

But, in the opinion of A. Caramazza this is not the crucial proof for mirror neurons existence, at least with the cognitive features that Gallese, Iacoboni and Rizzolatti assign to them. According to Caramazza, to value if human brain contains mirror neurons, may be useful a medical technique called fMRI adaptation. This technique allows to us to understand if a specific cerebral area is sensitive to the change of a property of a stimulus (e.g. the colour or the shape), or not sensitive to a similar change. The principle is that the repetition of a stimulus determines a less strong response of the nerve cells involved. Mirror neurons should be sensitive to a change of motor acts, not depending on the fact that the motor act is observed or performed. But this doesn't happen in the research carried out.

In fact, the motor acts taken into considerations usually have “target objects”, as a pen, a cup, etc., for which it is not clear if the neurons activation is about the properties of the target object or the movement in itself. From Caramazza's point of view there are two principal models or paradigms of the architecture of the mind in the brain: 1) the reductive-eliminativistic perspective, which claims that the whole cognition may be reduced to sensori-motor representations; 2) the non-reductive perspective, which claims that cognition is not at all reducible to sensori-motor representations. The most important point of

disagreement with Gallese and Rizzolatti too is about the interpretation of mirror neurons role. Their existence is consistent with a potential role in the action understanding, but till now we haven't had any important clinical demonstration that mirror neurons have really a functional role in the action understanding and in which way they could do that. In Caramazza's perspective there is a relevant gap between the original discovery of mirror neurons which show a selectivity in relation to motor acts and their involvement in human cognitive functions. Higher cognitive human functions cannot be usually reduced to simple relations just like the sound of breaking of a nut and associate it with its own motor act.

There are many examples in our everyday experience in which the same visual input (e.g., a yawn) can take on different meanings (fatigue, boredom, provocation, malaise) that can be understood only thanks to other background informations which are probably not available on a selective basis to the motor system. It's necessary to distinguish between the data which are consistent with the involvement of MNs in action understanding and the data which really show that mirror neurons have a crucial role in action understanding.

On the other side of the dispute MNs discovery and its interpretations, V. Gallese claims that mirror neurons certainly annoy those who look at neurosciences only as mere instruments of localization and validation of mental/psychological mechanisms deemed valid a priori. When neurosciences produce results that challenge or even refute these models, the first reaction is to deny the existence of these results. Gallese states that what Caramazza says about mirror neurons -beyond the inherent limitations of his recent work considered as the proof of the inexistence of mirror neurons in man- would be a clear example of this attitude.

During the last ten-fifteen years from MNs discovery in the macaque brain, several researches deeply changed both the traditional way of conceiving the relation between perception and action, and the role that perception and action have in the construction of social cognition. The discovery of the mechanism of motor resonance in MNs has shown that the motor system, far from being merely a muscle controller and a simple executor of commands coded elsewhere, is able to perform cognitive functions that have long been erroneously considered the prerogative of psychological processes and neural mechanisms of a purely associative kind.

The perception of action in the others, that is the understanding of the specific intentional aim in physical movements, which are not pure and neutral, appears to be a intrinsic mode of action, since it is rooted in the motor knowledge which underlies our own capacity to act. Classic cognitive sciences have hard problems, in Gallese's opinion, to accept this idea because, according to their traditional architecture of our mind, motor system cannot have *prima facie* (in principle) cognitive features.

The embodied mechanism in MNs would provide us, on the contrary, a more rich series of physical processes that underly social interactions, starting from filogenetical and ontogenetical processes. Our understanding of actions and motor intentions, made possible by MNS, puts in jeopardy the abstract mentalism or “mentalese” supported by many cognitive psychological models, starting from

the classic paradigm of the modularism supported by Fodor in his influent theory of mind.

The debate on the capacity to understand the others has been faced, in Gallese's opinion, in a bad way for many years, conceiving solely this capacity to "mindread" (this is the definition used by Alvin Goldman see "theory of mind" wikipedia), that is to attribute beliefs and desires to other people, in the light of a priori and abstract psychological theories, based on propositional attitudes. By this way, it is given a psychological explanation considered right a priori, but without explaining both real neural mechanisms and consequently psychological processes underlying social cognition.

For many years it has been told that when we understand the behavior of others, specific cerebral areas are activated, as the anterior cingulate cortex (ACC) and the temporo-parietal junction (TPJ), that were considered the cerebral places of an alleged module of the cognitive theory of mind. But this is false. In fact it has been demonstrated that these areas are activated even for tasks completely unrelated from the "mind-reading", such as attention or sexual excitation. In Gallese's opinion, cognitive theories about mind-reading are a sort of phrenology of the XXI century, considering also the fact that cognitive psychologists know few things or nothing about how the brain works. On the contrary, neuroscientists try to understand which is the neurophysiological mechanism that determines the activation of a certain cortical circuit during a task. The reality is that nobody till now know why ACC and TPJ sistematically activate even during mentalistic tasks, because we ignore again which is the neurophysiological mechanism that underlies their activation.

Caramazza claims that it is unlikely that MNs -if they really exist- can play a role in complex functions as empathy and language understanding or in explaining cognitive pathologies as autism. What philosophers, sociologists and people interested in marketing wrote in these last years about mirror neurons and their way of working and capacity to explain is often wrong.

Gallese doesn't agree with Caramazza both in denying MNs existence or in reducing so much their role. According to Gallese, we cannot prove that they don't exist, while we have many experimental data and scientific works in favour of their existence, obtained by different researching techniques as PET, Fmri, TMS, EEG, MEG and at last, at the beginning of April 2010, the planting of some electrodes in the brain of twenty-one patients of Iacoboni treated for epilepsy that would clearly show mirror neurons existence. In fact, according to the theory of MNS, MNs fired both for action and observation, and their excitement was directly recorded by the electrodes. The activity of 1177 MNs was recorded also in areas in which it was not yet assumed their presence.

The presence in human brain of a mechanism similar to MNS represents, in Gallese opinion, the most parsimonious unifying explanation for several behavioural and clinical data. Since it has proved that MNs exist and have an adaptive-evolutive role in birds and monkeys, it would be a non-sense or at least a strange thing in the evolutionary processes that such a MNS, which is so useful for the survival of many natural species, is absent in human being, whose evolutionary story is not longer than that of birds and monkeys.

But it is clear that the cognitive complexity of the human being leads us to wonder what mirror neurons and mirror neurons system can explain about human behaviours and what they couldn't explain. This is the reason why there is a wide interest and a fruitful debate among many researches in disciplines as philosophy, psychology, sociology, etc. about the interpretation of this discovery. Gallese is one of the most known fan and supporter of the so-called "cognitive neurosciences", whose aim would be that of looking for the neurophysiological bases of social cognition. He is convinced that the hard problem of how to explain intersubjectivity cannot be faced and solved uniquely either by philosophy or neuroscience or psychology, but it requires an interdisciplinary approach.

As I told in the Chapter 1.3 (*The epistemological dualism*), mind-body problem requires an interdisciplinary approach and the status and role of cognitive neuroscience is a good example of this multi-level approach to understand and explain in a better way complex phenomenona as consciousness, the self, intersubjectivity, etc. The idea of "neurophilosophy" supported and endorsed by P.&P. Churchland goes precisely in this direction. Gallese, Churchland and other (neuro)researchers also claim the need for a new common language which is able to put together in a harmonious way the different perspectives of research carried out by the new theories from neuroscience, philosophy, psychology, etc., to understand and explain cognitive complex phenomena.

Thanks to a better scientific disclosure, Gallese thinks that there is a more and more interest in mirror neurons discovery and cognitive neurosciences, generally speaking, because, by this way, people are really realizing that this is the best way to know who we are and how we function. Notably, MNs capture people attention because they show in a clear and simple way the crucial role of empathy in allowing us the mutual communication and understanding by emotional modalities rather than by sophisticated and abstract logic inferences and operations. From MNs discovery, it has begun to investigate in a scientific way intersubjectivity and its pathological alterations from a physical and neural perspective, very tied to the body. By this way it has been possible to reactivate and revitalize the philosophical tradition, above all the phenomenological view, in the direction of a phenomenology of perception notably deepened by M. Merleau-Ponty.

Nevertheless, Gallese himself is clear to state that he didn't claim and don't claim that MNs are able to explain everything there is to explain about social cognition. From his point of view they allow us to understand basic aspects of subjectivity both from a phylogenetic and ontogenetic point of view. And this understanding may have important repercussions on the understanding of the mechanisms underlying more sophisticated forms of social cognition. It is and empirical question (luckily not an article of faith!) to understand how far you can go using the mechanism of mirror neurons as key of reading for social cognition.

In the end Gallese doesn't agree with what the psychologist Paolo Legrenzi and the neuro-psychologist Carlo Umiltà state in their last book, "Neuro-manie", Il Mulino Publisher, Bologna 2009, that brain doesn't explain who we are. He doesn't agree with them because this statement and some others promote in our Italian

society, already historically skeptical and ignorant about science, a higher separation between scientific and humanistic culture. Certainly this book has received the greatest approval by the extra-scientific circles, as the enthusiastic review published on the Vatican Newspaper, "L'Osservatore romano". Gallese share with Legrenzi and Umiltà the alarm for the excessive use of neuroimaging techniques, but doesn't agree their criticism against cognitive neuroscience.

The neuroscientific explanation of a cognitive or behavioral feature cannot be reduced to a mere location, but it is really an explanation when it is able to identify the mechanisms that enable the activation of a specific brain circuit while performing a specific task. This is the important and distinctive contribution of cognitive neuroscience. Otherwise, there is the risk to do bad reasonings.

### **3.2. Are We Sure that MNs Exist? P. Pascolo's Opinion about Them (August 5, 2009)**

If the neurons that mirror the action (a goal action) are the same neurons performing the action, what happens in the case of contemporary execution of these two kinds of actions? A competition? Double circuitry? How a mirror system arises in an animal, for example a horse, when performing an "equivalent" act such as opening the door of the stable with the mouth?

(Paolo Pascolo, Extraordinary Professor of Industrial Biomechanics at the University of Udine)

Pascolo, anticipating in a study of 2008 the criticism of the research group of Caramazza about the existence or, at least, the cognitive role played by mirror neurons, criticized in a radical way the theories of Rizzolatti and his colleagues at the University of Parma about the existence and the role of mirror neurons. The first couple of questions that Pascolo takes into consideration are:

- i) if the neurons that mirror the action (a goal action) are the same neurons performing the action, what happens in the case of contemporary execution of these two kinds of actions? A competition? Double circuitry?
- ii) How a MNS arises in an animal, for example a horse, when performing an "equivalent" act such as opening the door of the stable with the mouth?

If we take in exam the many works from 1996 (the year of the official communication of mirror neurons existence in the monkey), we realize that "neuronal times" and "gesture times" are linked on a unique base of times which doesn't distinguish in a precise way the different phases in the execution of a gesture, that is facial expressions, intentions, beginning of the movement, contact with food, grabbing, etc. Therefore, the neural date checked in the monkey has an uncertain neural timing. In some cases the monkey was in delay while in others was ahead. Which is to say that monkey wasn't copying the gesture but was thinking in its own way: "I know what you are doing" or "if I am in your shoes I'll do it in this way" or "I already know what you are going to do", etc.

It often happens to anticipate with our own thought the gesture that we expect to see in the man we are observing. It is usual that a parent who is observing his child while he is understanding, often anticipates with his own thought or actions what he expects from him. So, it should be the case to read again the experiments of the Parma team in a way

to not take anything for granted. Pascolo underlies that in the whole literature on mirror neurons, it has been given too much for granted that mirror neurons are "seen" in the monkey brain. Experiments on the monkeys would have "certified" the existence of mirror neurons and several researchers focused on experiments about human being.

Till now it has assumed that the experiments were valid and consistent. Experiments on human beings are not invasive and direct. Therefore they have only a circumstantial nature. Many researchers have questioned the effectiveness of fMRI experiments to show the existence/non-existence of mirror neurons in human being. The Parma team would not have instrumented in the right way the animal or the investigator with suitable sensors to perceive acts of motion, facial expressions, etc. The "anticipatory maneuvers", according to Pascolo, could be explained without relating to the MNS. A simple anticipatory maneuver could be the following: when I'm going to grab an object wringing the hands forward, I will execute a retreat in advance of the basin. I will counteract the effects induced by the inertial movement forward and prepare the body system in balance, to accommodate the load.

In practice, I examine the hypothetical charge that will manage and implement strategies with the necessary experience (Pastore et al., Chaos, *Journal of Biomechanics*). You may apply your experience also with the Inno of Mameli: if you know the words you may sing in chorus, anticipate, delay, counterpoint, etc., while if you don't know the words you will stutter or sing badly and in delay. Anticipation would be nothing but experience. When you are used to have many hours of training, as in the case of a boxer, and you got many matches, you are able to anticipate the opponent's blows. On the contrary, when I use a certain neural circuit to examine a gesture and to execute it, I have to consider a necessary time for making possible the mirroring function of the neuron. This means delay, at least 130-150 milliseconds because there must be some different and necessary physiological processes: interpretation, transmission to motor-neurons and muscle involvement.

It's clear to everybody that when we are used to move and behave in a certain way, playing football or tennis, etc., we are consequently used to anticipate our movements and behaviors according to an experiential map in our brain and body. To Pascolo there are not absolute scientific evidences for the existence of MNS in monkeys and humans since the experimental subject is motionless under fMRI and the data refer to the thermal effect produced by visual stimulations and it is detectable in a considerable number of neurons. In conclusion Pascolo states that MNS theory should be revised. If we are not sure about the existence of MNS, we should also revise some experimental protocols for rehabilitation and the interpretation of autism.

### **3.3. The Critique to Cognitive Role of Mirror Neurons: P. Jacob's Theory (June 8, 2009)**

"Human ability to represent psychological states (beliefs, intentions, desires, emotions) and to attribute them to others (so-called mindreading) goes beyond the mechanism of mirror neurons. Consequently, also the idea that autism stems from lack of mirror neurons is wrong."

(*Pierre Jacob*, Philosopher of Mind and Cognitive Scientist, currently President of the European Society of Philosophy and Psychology and Director of the Jean Nicod Institute in Paris).

Cognitive neuroscientists, particularly those ones of Parma University, the team headed by Rizzolatti, starting from MNs discovery in the premotor ventral cortex of the macaque monkeys, stated the existence of such a MNS in man thanks to neuroimaging, but what they call MNS includes an important area (the superior temporal sulcus) which doesn't contain for itself MNs. P. Jacob underlies the fact that MNs are not able to explain all the cognitive functions which are necessary to make possible the understanding of intentions in other people and the complex phenomenon of empathy. If MNs fire in the brain of a monkey or of a man who is watching his conspecific while he is grasping an object, what the activity of MNs generates in the observer would be only a mental repetition of the act of a man who is acting. But there is a fundamental gap between the mental repetition of the act of a man who is grabbing a cup of coffee and the cognitive capacity to know if the agent is going to do it for drinking the cup of coffee or for giving it to somebody, or for putting it on the table or for throwing it out the window, and so on. Repeating or simulating in a mental way the motor act of an agent is not clearly sufficient to understand the intention behind that action, for example, giving the cup to our dear friend.

It is not even sure that this mental repeating or simulating is necessary for understanding the intention. Moreover, we are speaking of transitive acts, that is grasping, squashing, etc., for which the aim is considered essential for the act in itself. On the contrary, probably nobody would say that while we are watching the act of grasping we may or should prove a feeling of empathy in relation to the agent to whom our act is directed. It seems that empathy is important and perhaps necessary only to understand and answer to affective internal states, (technically qualia), of the other people, that is pain, happiness, etc., while would be not important and necessary to understand motor acts. Naturally affective internal states may be manifested by behaviours, but they are not, properly speaking, acts.

Human capacity of representing psychological states (beliefs, intentions, desires, emotions) and attributing to others (so called mindreading) go beyond the mechanism of mirror neurons. Therefore, the idea that autism arises from a deficit of mirror neurons is wrong. It is possible that autistic people have a deficit of mirror neurons (if they really exist), but also in this way it is certainly possible that the deficit of mirror neurons comes from a deficit in the capacity of mindreading, and not vice versa. Perhaps it is necessary to represent and understand the intention of an agent to make mirror neurons active. In this case the activity of mirror neurons doesn't generate the understanding of the intention of an agent. The causal chain goes in the opposite direction: first, you understand the intention of the agent, then MNs internally imitate (or mentally reproduce) the observed act.

### 3.4. A Phenomenological Interpretation of MNs: L. Boella's View

L. Boella, Professor of Moral Philosophy at the University of Milan, states that the discovery of mirror

neurons has certainly contributed to the actual and popular success of neurosciences. From Boella's point of view this success has been favoured by the special evidence and simplicity of the result of this discovery. By this way MNs discovery has spread through many disciplines, being interpreted beyond the specific context of the experimental research.

It is particularly significant that this discovery interested the philosophers. In fact it focuses on a central point of the contemporary thought, the belief that the inter-subjective relationship and the acknowledgement of the other are essential for the individual and society. Therefore, it is happened a particular convergence with the French phenomenology represented above all by Maurice Merleau-Ponty, in the direction of a new interpretation of perception, considered not only as an assembly of sensible data (visual, auditive, ecc.), but as a real dialogue with the external world (things and people) in which the body is the protagonist.

In fact it shows, by its own movement, intentions and preferences, and in the same time knows the world discovering it. In France researchers and philosophers (as Jean Luc Petit) have uploaded the scientific references to Merleau-Ponty, working on the link between perception and action, using experimental data as those ones produced by experiments on MNs. The point of meeting between MNs discovery and philosophy concerns *in primis* the visual-motor role of MNs and their role in the perception of finalized actions. Recently (G. Hickok, 2009), the non cognitive mechanism of association between actor and spectator ("I know what you are doing"), concerning mirror neurons, has been tied with the "motor theory of knowledge", which was very common among psychologists of the beginning of 20<sup>th</sup> century.

For sure the experiments on MNS has become an obliged reference to every philosopher who makes a particular research on empathy. The "rediscovery of empathy" (Stueber, 2006, Goldman, 2006) is due to the interpretation of MNS as an "enlarged empathy". In fact, in this MNS researchers have seen the neurobiological base for a constellation of private states, notably qualia but also behavioural attitudes, from sympathy to the comprehension of the other people till the care. Having studied very much many philosophers and psychologists who took into consideration empathy, Boella is very impressed and sceptical about some forced implications between MNS discovery and the theories on empathy taken on by some philosophers and psychologists (e.g., Theodor Lipps) who played an important role in the reflection on empathy, but often offered a poor version of it.

The philosophical reflection on empathy achieves the maximum deepening in the context of the first expression of the phenomenological movement, that is between 1910 and the first half of '20es, thanks to Husserl, Scheler and Stern. The most important thing is that empathy is not longer considered as a moral feeling (piety, compassion, sympathy), but as an autonomous experience. Before sharing the pain of somebody else, I must recognise the concept of "somebody else", understanding that he has feelings, thoughts, volitions, just like me, even if he is different from me. So, first we have what is called "the discovery of the other", the acknowledgement of his existence, his "being part of the world" in which I'm living. Only on this base, it has a meaning for us to speak

of a passage of feelings from me to another, a sharing of feelings as solidarity. This is a fundamental premise because it is a crucial point for the philosophical thought on empathy.

In fact, if we look at contemporary philosophy we see an interrupted or uncompleted way. The reason is in the fact that phenomenologists separated the empathic experience from its bodily origin (here the reference is above all the last thought of Husserl and Heidegger, and all their students). Notably, they directed their attention to the “Lebenswelt”, the “world of qualia”, that is the natural, social, linguistic and cultural horizon in which we are in relation with the others in an unconscious way, losing by this way every interest for the biologic-organic roots of empathy.

## 4. Some Philosophical and Psychological Implications of MNS Working

### 4.1. MNS and Intention Understanding

Actions, according to standard philosophical wisdom, are in some way conceptually tied to intentions. An action is something an agent intentionally does.<sup>13</sup> So, to recognize that some behavior (hand movement) is an action (waving hello), one must understand the intention with which the agent acts (to greet you). An intention is standardly regarded as a mental state (Davidson, 2001; Grice, 1972; Searle, 1983). As such, intentions are something distinct from the behavioral sequence and are not directly detectable from the behavioral sequence. I endorse the view that intentions are mental states and that explaining intentional action involves attributing some sort of mental state.<sup>14</sup>

Strictly and broadly congruent mirror MNS fire in response to particular details (that it is a whole-handed grasp, or that it is an eating-related grasp) about a behavior, but there is a metaphysical and epistemological gap between this mirror neuron activation and the mental state attribution required for action understanding. On my view, the activation of broadly and strictly congruent mirror neurons is like yawning in response to observing another’s yawn. My yawning reflex does not constitute understanding or in any way imply that I understand that you are bored. Understanding that you are bored requires recognizing that you are in a particular mental state, and my yawning response does not constitute that recognition.

Likewise, the automatic resonance of my MNS does not constitute recognizing your mental states. Broadly and strictly congruent mirror neuron activation in an observer entails nothing about whether the observer attributes an

intentional mental state to the actor. The same is true for logically related MNS. Neural firing in expectation of an event in a behavioral sequence does not constitute understanding another’s mental states. These different kinds of mirror neuron activity may all be partially causally related to action understanding, with each kind of mirror neuron providing different kinds of information about the observed behavior, but none of the kinds of mirror neuron activity, individually or collectively, constitute action understanding. Analogously, my yawning in response to your yawn may play a role in my inference that you are bored, but the yawning plays a metaphysically and epistemologically indirect role.

Moreover, no kind of mirror MNS activity is necessary or sufficient for genuine intention understanding. It is not necessary because we can infer an intention without mirror neuron firing. For example, suppose I tell you that Johnny went to the store and bought graham crackers, chocolate and marshmallows. From this you infer that Johnny intends to make s’mores. Although I have not done any brain scanning to test whether your mirror neurons would be activated in this scenario, I find it unlikely that they would be. What exactly would the mirror neurons mirror? The typical motor mirror neuron activations are for hand, foot and mouth movements, but none of that information plays a role in the story I describe, and yet you are still able to infer the intention.

Furthermore, mirroring is not sufficient for intention understanding. Automatic neural resonance and anticipatory neural activation are not sufficient for understanding an intention. The automatic motor resonance, which I likened to yawning in response to observing a yawn, is not sufficient for understanding an intention, a mental representation. And neural firing in expectation of an event in a behavioral sequence is not sufficient for intention understanding, either.

### 4.2. MNS and Goal Understanding

MNS activity, on my view, is more closely related to understanding goal-directed behavior than intentional actions. Understanding a target’s goal-directed behavior amounts to understanding the target’s orientation toward some thing in the world, which requires various motor and sensory representations. In contrast, actions transcend mere behaviors. An action is something an *agent* does *intentionally*. Understanding a target’s actions requires not only understanding the target’s orientation toward some thing, state or event, but also understanding how the target represents her orientation toward that thing, state or event. That is, understanding an action requires understanding the target’s mental representations (Davidson, 2001; Searle, 1983). It is standard in philosophy to distinguish behavior, the understanding of which requires motor and sensory representations, and action, the understanding of which additionally requires mental representations.

My insistence on making this distinction between goal-directed behaviors and actions may seem like a mere terminological squabble, but this is not the case. I do not care what we *call* these two categories so long as we keep them distinct. The reason I am at pains to distinguish the two categories now is that, as will become clear in the next section, in the discussion of the cognitive importance of mirror neurons this distinction is often neglected. If

<sup>13</sup> In the action theory literature, theorists distinguish three senses of intentional action: acting *with* a certain intention, acting intentionally, and intention-for-the future. There have been various attempts to give a unified account these senses of intentional action (Anscombe, 1957; Davidson, 2001).

<sup>14</sup> There is a debate over exactly what kind of mental state an intention is. I shall sidestep this debate here. In endorsing the intention-as-mental-state view, I am rejecting a host of non-standard views, inspired by Anscombe’s account, that understand intention in terms of goal-oriented behavior (Moran and Stone, 2008; Thompson, 2008). These views hold that intending to do A is not a mental state; it is a form of being in progress toward some goal. Thus, on these views explanation of intentional action need not invoke any psychological terms. These non-standard views will come up again later in this chapter.

mirror neuron activity constituted action understanding in the full philosophical sense of the term, then mirror neuron activity would constitute intention understanding, and this would have important implications for how mindreading is accomplished. But if MNs simply activate for goal-directed behavior, then the relationship between mirror neurons and intention understanding is less direct and less clear, and the relationship between mirror neurons and mindreading is even more tenuous. I shall argue that, contrary to the many bold claims about mirror neurons and social cognition, the latter claim is true.

On my account, MNs are more closely related to understanding goals than intentions. However, MNs are still only tenuously related to goal understanding. Mirror neurons do not constitute, and are neither necessary nor sufficient for, understanding goal-directed behavior. MNs can be causally relevant to goal understanding. Importantly, though, they are not the only relevant areas of the brain for understanding goal-directed behavior. In addition to MNs, neurons in superior temporal sulcus (STS), canonical neurons, and non-motor perceptual cues play key roles in understanding goal-directed behavior.

Neurons in the STS have the same perceptual properties as mirror neurons but lack first-person motor properties. In other words, these neurons fire only when observing the target's goal-directed behavior, never simply when the subject acts. The STS is an area where others' behaviors are visually processed and has long been recognized as part of the neural circuitry underlying the perception of others' behaviors (Gazzaniga, et al., 2009, p. 549).

Canonical neurons are the inverse of STS neurons; they have the same motor properties as mirror neurons but differ in their perceptual properties. These neurons fire when the subject grasps objects and when the subject sees a graspable object, but not when a target grasps an object. Canonical neurons are thought to process, in the first-person case, one's own motor movements toward an object and, in the third-person case, the potential for behavior directed toward the observed object (Gazzaniga, et al., 2009, p. 550).

Non-motor perceptual cues are also relevant to understanding goal-directed behavior. In an influential study on mirror neurons, researchers found that mirror neurons in monkeys preferentially responded to grasping-to-eat over grasping-to-place behaviors even when these behaviors were motorically very similar (Fogassi, et al., 2005). For our purpose, the important feature of this study is that two factors helped the monkeys discriminate between grasping for eating and grasping for placing: whether the object grasped is food and whether a container is present in the context of the perceived action (Jacob, 2008). Both of these factors are purely perceptual cues. Importantly, purely perceptual cues do not themselves cause mirror neuron activity. If shown a picture of a container and a piece of food, one's mirror neurons would not fire. Only observing or performing motor acts causes mirror neuron activity. And yet perceptual cues are relevant factors in recognizing that some movement is a goal-directed behavior. My hypothesis is that perceptual cues modulate mirror neuron responses. That is, when observing some movement, perceptual cues aid in determining whether the movement is a goal-directed behavior. Thus, even if we narrow our focus from actions to goal-directed behavior, mirror

neurons are still only a limited part of the story. Non-motor perceptual cues and STS and canonical neurons are proof that mirror neuron activity does not constitute, nor is it sufficient for, understanding goal-directed behavior.

On my account, MNs causally contribute to understanding goal-directed behaviors, and this may play a role in understanding intentional actions. I regard mirror neuron activity as a contributory cause of understanding goal-directed behavior. In other words, MNs are neither necessary nor sufficient for goal understanding, yet they still causally contribute to understanding goal-directed behavior. Mirror neuron activity is not sufficient for understanding goal-directed behavior because determining whether some behavior is goal-directed depends on non-motor perceptual cues and non-mirror neuron areas of the brain. Furthermore, although mirror neuron activity may *in fact* be a mechanism we use to understand goal-directed behavior, it is certainly not *logically* necessary for understanding goal-directed behavior. We can imagine creatures (or computers!) that can recognize goal-directed behavior and yet lack MNs. I doubt that mirror neuron activity is even nomologically necessary for goal understanding, but my arguments do not hinge on this claim.

### 4.3. MN Firing and Action Understanding

Empirical evidence supports the hypothesis about the relation between mirror neuron firing and action understanding. First, mirror neurons are sensitive to the mode of presentation of actions. For example, monkeys' MNs do not fire when watching a familiar behavior on a video monitor (Ferrari et al., 2003; Keysers and Perrett, 2004) despite the fact that there is evidence that the monkey understands the behavior on the monitor.<sup>15</sup> This casts doubt on the claim that mirror neurons constitute or *directly* cause action understanding (Jellema et al., 2000).

Second, humans' mirror neuron activity is insensitive to the difficulty of interpreting an action. Brass, et al. (2007) hypothesize that our remarkable capacity to flexibly interpret observed behaviors as intentional actions is mediated not by the mirror neuron system, but an inferential interpretive system located in the STS and anterior frontomedian cortex (aFMC), areas independently associated with perception of social stimuli, mentalizing, and action understanding. This study tests that hypothesis by having subjects, while in an fMRI machine, watch three short videos in which an actor operates a light switch with her knee.

The three videos demonstrate the actor operating a light switch with her knee in a plausible context (both of the actor's hands are fully occupied), implausible context (actor uses two hands to hold a small, lightweight item), and no context (actor's hands are unoccupied). The subjects are required to come up with a rationale for each case. Experimenters found that activation in the STS and aFMC activated to a level corresponding to the difficulty in ascribing a rationale to the actor. In other words,

<sup>15</sup> For example, in their foundational study on chimpanzee's theory of mind abilities, Premack and Woodruff (1978) had a chimpanzee watch a video of a human attempting to do something, and they stopped the video before the human achieved the goal. The chimpanzee had to choose from a set of photographs depicting various outcomes, and in 18 of 20 cases the chimpanzee chose the correct photograph, i.e., the photograph that depicted the human completing the goal. This shows that monkeys can understand behaviors displayed on video monitors.

attributing an intention to the actor in the plausible context elicited the lowest activation of the STS and aFMC, and attributing an intention to the actor in the no-context scene elicited the highest activation of the STS and aFMC.

The more difficult it is to ascribe an intention to the actor, the more strongly these areas activate. Subjects' mirror neuron activity, in contrast, was the same for each condition. Mirror neuron activity does not differentiate between harder-to-interpret actions and easy-to-interpret actions. This undermines the idea that mirror neurons constitute or directly cause genuine action understanding. Genuine intentional action understanding is mediated by the STS and aFMC, an inferential interpretive system.

A third source of evidence for my account comes from studies on social non-human animals. Many non-human animals understand goal-directed behavior, and as new research continues to reveal, many non-human animal species have mirror neuron systems. Scientists argue that various monkey, dolphin, and elephant species have rudimentary mirror neuron systems (Blakeslee, 2006). The animals in which mirror neurons are present are capable of understanding goal-directed behaviors, as evidenced by various studies on the social behavior of these animal species. For example, classic behavioral experiments in cognitive ethology, which test monkeys' social cognitive skills, reveal clear limitations on monkeys' social cognitive abilities (Povinelli and Vonk, 2003), but they also show that monkeys are not merely behavior-reading. Experiments in cognitive ethology reveal that some monkeys are sensitive to the gaze direction of conspecifics and humans, follow others' gazes to out-of-view objects, and take into account opaque barriers (Tomasello, *et al.*, 2003). They can also adapt their food retrieval strategy based on whether a dominant competitor can or has seen the food location, and they can even manipulate whether a competitor can see them to gain strategic advantage (Hare, *et al.*, 2000; Hare, *et al.*, 2001; Hare, *et al.*, 2006).

These experiments indicate that monkeys are capable of understanding goal-directed behavior, but there is no unequivocal evidence that all these species are capable of genuine intention understanding. My account offers a unified explanation of these findings. MNs, in conjunction with other motor and perceptual mechanisms, work to detect goal-directed behavior, which the behavioral studies have shown these non-human animals are capable of detecting. If, contrary to my account, mirror neuron activity were to constitute action understanding, and thereby genuine intention understanding, then either we would have to accept the *prima facie* dubious conclusion that all these animal species have low-level mindreading abilities similar to ours *in virtue of* the similarity between our mirror neuron systems, or we would have to find some substantial, non-ad hoc difference between their mirror neuron systems and ours. I think neither of these options is particularly compelling.<sup>16</sup>

<sup>16</sup> It may be the case that some kinds of monkeys are capable of mentalistically understanding behavior. That is, some species may be capable of genuine intention understanding. If that is the case, it is not in virtue of their mirror neurons on my view. For all monkey species have mirror neurons, but social cognitive skills varies across species. Chimpanzees, for example, exhibit more sophisticated social cognitive behaviors than macaques. As a matter of fact, this is one of the reasons why research protocols allow single cell experiments on macaques but not on chimpanzees. Thus, even if some monkeys understand intentional attitudes, mirror neurons are not the relevant causal factor.

#### 4.4. The Tea Party Experiment

Let's look at a particular study to see how my account explains the results. This example is from a foundational study on motor mirror neurons in humans referred to as the Tea Party experiment (Iacoboni, *et al.*, 2005). Subjects observe the following scenes: *context*, *action*, and *intention*. The *context* scene contains only objects, e.g., a teapot, a cup, a plate with cookies. There are two kinds of context scenes. In one *context* scene, the objects are arranged neatly, suggesting that someone is going to have tea. In the other *context* scene are crumbs, a dirty napkin, and a tipped-over cup, suggesting that someone has already had tea. In the *action* scene, subjects observe a hand grasping a cup without any contextual cues. The *intention* scene combines the *context* and *action* scenes, and subjects observe a hand grasping a cup in the neat or messy context.<sup>17</sup>

The researchers found higher activation in MN areas while subjects observed the *intention* scene embedded in the clean and dirty contexts, compared to observing the *action* scene. They also found higher activity in mirror neuron areas while subjects observed the *intention* scene with the context that suggested drinking, compared to the context that suggested cleaning up. In addition, half of the participants in this study were instructed to pay attention to the intention displayed by the behavior they were observing, while the other half were not told anything about intentions. The researchers found no difference in MN activation between the participants in each group, but in the debriefing session all participants were able to report accurately the intentions associated with each version of the *intention* scene.

What should we make of the data from the Tea Party experiment? There is some mirror neuron activity in subjects observing the *action* scene, but there is higher activity when they are observing the *intention* scene, especially the neat version of the *intention* scene. On my view, the mirror neuron activity in the *action* scene is due to the activation of strictly and broadly congruent MNs, which motorically resonate in response to the observed features of the behavior, providing particular bits of information about the observed behaviors.

The increased activity during the *intention* scene is due to the additional activation of logically related mirror neurons, which function as an anticipatory or predictive mechanism. It has been shown that mirror neurons activate more strongly when observing familiar behaviors (Calvo-Merino, *et al.*, 2006). When a subject observes a familiar behavioral sequence, his logically related mirror neurons fire in expectation of a certain behavioral event. In the experiment, logically related mirror neurons do not activate for the *context* scene because there is no goal-directed behavior to predict. The *context* scene contains only objects. Logically related MNs do not activate for the *action* scene because, again, there is nothing to predict. The *action* scene shows a hand grasping a cup, and there is no information that indicates any goal beyond grasping the cup. The increased activation of logically related mirror neurons for the *intention* scene indicates an expectation of further unobserved behavior, drinking (in

<sup>17</sup> For reasons that should be clear by now, I think the titles *action* and *intention* are misleading.

the neat version) or cleaning, perhaps (in the messy version).<sup>18</sup>

This mirror neuron activity, in conjunction with the perceptual cues (e.g., crumbs and dirty napkins), neurons in the STS and canonical neurons, cause one to recognize, and anticipate the completion of, a goal-directed behavior. This expectation or prediction may sometimes be involved with intention understanding in the sense that it may provide information relevant in inferring intentions, but it is not constitutive of intention understanding.

Finally, what should we make of the post-experiment debriefing about the intentions associated with the neat and messy versions of the *intention* scene? In light of the Brass, et al. study explained above, I think we have pretty good evidence that mirror neuron activation does not reflect understanding intentions associated with each scene. I will discuss this fully in the next section. For now, I shall just say that I think the post-experiment debriefing is uninformative with regard to the neural correlates of intention understanding. There are many possible explanations of this feature of the Tea Party experiment, and the hypothesis that this feature proves that mirror neuron activity constitutes intention understanding is plausible only if you already believe that MNs underlie intention understanding.

## 5. The Theoretical Concordance of MNs Discovery with the Supervenience Reductive Theory of Mind by J. Kim

### 5.1. Psychophysical Supervenience and Mental Causation (MC)

Now I'm going to show the theoretical concordance of this discovery with a reductive theory of mind, and particularly with the supervenience reductive theory of mind claimed by Kim (1993, 1996, 1998, 2005).

Indeed, it seems to me that the way in which the so-called "mirror system" works is theoretically compatible with some crucial concepts and principles in the metaphysics of mind by Kim. Notably, they are: 1) the concept of reductive psychophysical supervenience, according to which a mental property is realized by a species-specific physical/neural property, making use of a functional model of reduction; 2) the pre-emption of a physical cause on a mental cause and the redundancy and unintelligibility of mental causation; 3) the principle of physical causal closure, according to which there are causes in a genuine sense always in the physical domain; 4) the multi-layered metaphysical model of the world, which distinguishes between ontological "levels" (micro/macro properties) and theoretical/conceptual "orders" (physical, mental, social, etc.).

1) Now, let's start with order considering these points starting from the concept of "psychophysical

supervenience principle", canonically formulated by Kim: «The mental supervenes on the physical in that any two things (objects, events, organisms, persons etc.) exactly alike in all physical properties cannot differ in respect of mental properties» (Kim, 1996, p. 10).

We could say in very synthetic terms: no mental difference without physical difference. It's important to stress that this principle doesn't state that things psychologically indiscernible must be therefore alike in every physical character, but only the converse thesis. So, psychophysical supervenience only claims in a necessary way that two or more creatures can't be psychologically different, being yet physically identical (in a negative form), or that physically identical creatures must be psychologically identical too (in a positive form). We can recognize three basic principles which found the canonical concept of psychophysical supervenience, as it was officially introduced in philosophy of mind by Davidson, in his "Mental Events" (1970).

(i) *Covariance* (only apparently asymmetrical) of properties: if two things or individuals are indiscernible in relation to their base physical properties, they will be indiscernible in relation to their mental or higher-level properties too, while it's not necessarily worth vice versa.

(ii) *Dependence*: supervenient properties depend or are determined on/by basic-physical properties.

(iii) *Irreducibility*: supervenient mental properties are not reducible (in the canonical version of supervenience) to their physical base properties (from a nomological/explicative point of view).

Leaving aside the long and complex history of this concept and starting from the last interpretation of Kim (1998) in a reductive sense, it is formulated in different ways, as the following two:

[Indiscernibility definition of mind-body supervenience] Mental properties *supervene* on physical properties, in that necessarily any two things (in the same or different possible worlds) indiscernible in all physical properties are indiscernible in mental respects (Kim, 1998, p. 10)

[Mind-body supervenience in relation with time] Mental properties supervene on physical properties in the sense that if something instantiates any mental property *M* at *t*, there is a physical base property *P* such that the thing has *P* at *t*, and necessarily anything with *P* at a time has *M* at that time (Kim, 1998, p. 39)

They mean that a physical base property ("P") is necessarily sufficient for the supervenient mental property ("M") because supervenient properties depend or are determined on/by their subvenient species-specific properties and there's an ontological identity between them, being instantiated in the same time *t*.

It seems to me that these distinctive features of the concept of psychophysical supervenience are fully consistent in a theoretical way with the functioning of mirror neurons and the "mirror system", according to which perceptual and cognitive processes are realized on the same neural circuitry of motor processes, depending on their way of playing. By this way, a will or an intention may be read as a motor disposition realized on its own neural circuitry.

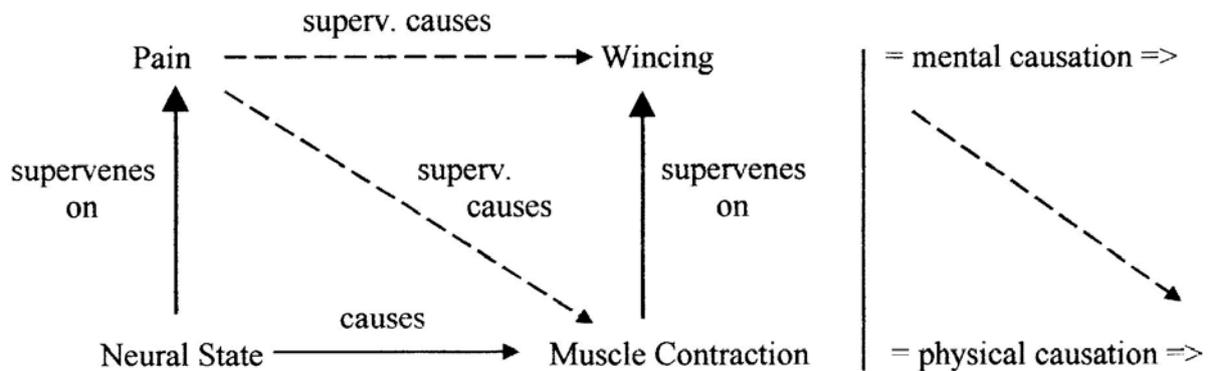
2) Hence, it's clear that a genuine causal action or causation takes place first at a physical level, in the activation of neural circuits that realized the motor processes, configuring the same intentionality as a

<sup>18</sup> The higher activation while observing the neat version of the *intention* scene may reflect the fact that drinking is pretty clearly the goal of a hand reaching for a cup in a scene set up as a tea party, whereas the goal of a grasping motion in the scene with crumbs and dirty napkins is more ambiguous. The scene does not unequivocally suggest cleaning. Perhaps the hand is reaching for crumbs to eat. Another possibility is that drinking is a more basic action than cleaning, and logically related mirror neurons fire more in response to basic actions. Either way, these results are compatible with my explanation of the results of the study.

preparation for action. By this way, as it says Kim himself (1998), mental causation comes to be redundant because physical causation is necessary and sufficient. In the same way Kim's functional model of reduction, according to which the gene is the role played by DNA and the lightning is the function performed by the electric discharge, is part of a reductive explanation of many behaviours that we would erroneously consider as totally

cognitive. Instead, in light of these discovery, studies and experiments, many cognitive behaviours seem to be much "embedded" and involuntary, in spite of appearance (as the awareness processes, e.g. decision making and will, namely described by Libet, 2004).

Let's see the picture in which Kim (1996, p. 51) distinguishes supervenience from physical causation and from mental or supervenient causation.



**Figure 1.** It's first the neural state, according to supervenience causation model, to cause, on one side, a certain neuro-physiological event, the muscle contraction, on which wincing supervenes, while, on the other, to constitute the supervenience base of pain, which causes, on a mental level, wincing

Generally speaking, this model considers higher order properties as necessarily dependent or supervenient on their realizers at a physical deeper level, and considers mental causal processes as necessarily dependent or supervenient on those at a physical deeper level, which is the only genuine causal level.

3) The reason is that physical domain is causally closed: indeed there's no mental or non-physical cause for itself, that is without a physical base of realization. Kim (1998) describes by these words his crucial "principle of causal closure": if you pick any physical event and trace out its causal ancestry and posterity, that will never take you outside the physical domain. That is, no causal chain will ever cross the boundary between the physical and the non-physical. (Kim, 1998, p. 40) Thus, an intrinsic mental causation cannot exist and it turns out to be completely redundant and unintelligible. But this doesn't mean automatically its exclusion, because its lack would make arise invincible explicative problems for understanding and explaining our cognitive faculties as language, reasoning, memory, etc.

From these considerations it the increasingly crucial role of the body and its physical laws appears to describe ontologically even the more abstract cognitive processes, following the idea of so-called "embodied cognition". But, as it is well known in the tradition of analytic philosophy, reasons are not causes. So if causes are physical descriptions, reasons and purposes may well cross the narrow domain of the physical world, and thus allow other more suitable explicative theoretical domains, such as psychology, sociology, etc.

### 5.2. MC between Level and Order Hierarchy

4) Finally, I consider Kim's distinction between "levels" (micro/macro) and "orders" (physical/mental), drawn above all in the third chapter of *Mind in a Physical World*, very interesting and important from a logical and metaphysical point of view, towards a reductive theory of mind. For it would deserve an adequate analysis and reflection both for its ontological and causal outcomes.

Let's get to the heart of the matter, following the analytic way to argue notably used by Kim.

His starting point, accepted by the majority of the actual philosophers of mind, is that mental properties or, generally speaking, second-order properties, are realized on physical/neural properties, as their base of supervenience. This realization relation should make the multilayered structure of levels arising. But now, which is the relation between level hierarchy and order hierarchy? Kim stresses the fact that «both second-order properties and their first-order realizers are properties of the same entities and systems» (Ib., p. 82). Which means to say that order hierarchy takes place on a same level, being first/second/n-order properties nothing but properties of the very same object or system on a certain micro/macro level. So, making use of Kim's significant words, «when we talk of second-order properties and their realizers, there's no movement downward, or upward, in the hierarchy of entities and their properties ordered by the micro-macro relation» (Ib., p. 91). And again, to differ order with level series: «the series created by the second-order/realizer relation does not track the ordered series of micro-macro levels; it stays entirely within a single level in the micro-macro hierarchy» (Ib., p. 82).

I'm just insisting on this distinction between order hierarchy and level hierarchy because I'm convinced that it is much more than a terminological distinction. Indeed it seems undoubtedly crucial at least to award causal efficacy to some properties. We may observe, for example, that order-properties within a supervenient progression from first (physical) properties to second (mental) properties, as C/fibers exciting property and the feeling of pain, are all properties applying to entities at a single micro-macro level with no further injection of causal powers at the higher orders.

On the contrary, «spin, charm, and such are properties of elementary particles, and they have no application to atoms, molecules, and higher objects in the micro-macro hierarchy; transparency and inflammability are properties of aggregates of molecules, and they have no place for

atoms or more basic particles». (Ib., p. 83) In the same way consciousness and intentionality are properties of biological organisms, or at least of neural systems, and they have no application to entities which are micro in relation to them. If this is right, we should correctly speak of first/second/n-order properties within a metaphysical hierarchy of orders only for a same object or system, and it is about this order hierarchy we should imagine and recognize the relation of psychophysical supervenience.

It is well known, then, that this logical and metaphysical relation of dependence and determination of mental properties on physical properties entails, according to Kim (1998), the idea of reducibility and causal inefficacy of mentality, whose apparent causal powers would be inherited from its physical base. This idea, from my point of view, clearly leads to a reductive way of seeing every mental process as it has shown by MNS working. Viceversa, micro-macro level hierarchy does not concern properties of a same object, but different properties for different objects, depending on their complexity in micro-macro progression. Within this progression you may speak of supervenience too, but only in a mereological sense, that is the “micro-based property” (as a “structural property” in David Armstrong) on its micro-constituents, such as, for example, a water molecule mereologically supervenes on two hydrogen atoms and one oxygen atom.

But no micro-constituent, none of water proper atoms, which constitute it, has already the causal powers it represents. Following Kim’s arguing, «H<sub>2</sub>O molecules have causal powers that no oxygen or hydrogen atoms have» (Ib., p. 85). In the same way, «a neural assembly consisting of many thousands of neurons will have properties whose causal powers go beyond the causal powers of the properties of its constituents neurons, or subassemblies, and human beings have causal powers that none of our individual organs have» (Ib., p. 85). In this light Kim’s functional model of reduction seems to be consistent with the idea of “emergence” of new causal powers climbing up level hierarchy, from micro-constituents level (such as atoms, neurons, etc.) through higher level properties (such as organs properties, apparatus properties, till human beings properties), but only if these complex properties are micro-based properties. By this way, we might claim proper causal powers for emergent complex properties, such as cognitive faculties and consciousness in human beings, in spite of their reductive and functional explanation via realization.

So, macro-properties in level hierarchy, in a different way from second-order properties in order hierarchy, can, and in general do, have their own causal powers, which go beyond the causal powers of their micro-constituents. Here it is the importance of the distinction between orders and levels, whose target is not only to make clear our language, but also to clarify which properties have proper causal powers and which ones has not. Now, if macro-properties could have, and generally have, their own distinctive causal powers, in addition to their micro-constituents properties, then we should recognize the possibility of a downward causation at least from macro-properties to their own micro-constituents properties. If we consider consciousness as a macro-property emergent or mereologically supervened on several micro-based

properties (just like the majority of philosophers of mind), such as organs and apparatus properties, we should conclude that it could have new genuine causal powers able to influence in a causal way lower level systems and constituents.

A key factor for attributing causal efficacy to the macro-properties is their temporal relation to micro-properties. Important experiments of the physiologist B. Libet (2005) on “time of consciousness”, show how the “timing” in the activation of brain processes, on one hand, and processes and/or “mental” causes (as will, decision making processes and conscious phenomena, generally speaking), on the other, is essential to maintain a reductive theory of mind. Coming back to Kim and mental properties (1998), the macro-properties of complex systems can have, and generally have, their own distinctive causal powers that go beyond the causal powers of their micro-constituents. It’s Kim himself to say it explicitly: «an assembly of neurons consisting of many thousands of neurons whose properties have causal powers beyond the causal powers of the properties of neurons that constitute, or sub-assemblies, and humans have causal powers that none of our own individual bodies» (Kim, 1998, p.85). Among the macro-properties Kim includes undoubtedly intentionality and consciousness, attributing special causal powers to them, because they are located on a higher ontological level: «consciousness and intentionality are properties of biological organisms» (Kim, 1998, p. 83).

About consciousness, intentionality and complex mental phenomena, the mirror neurons discovery, since it shows a close neuro-physiological link between motor processes and cognitive functions as perception, vision, etc, thus it confirms, in my opinion, the fact that our descriptive orders for a behaviour, such as intentionality, will, etc.. lie on the same physical level of the implementation of that behaviour, often even before our awareness, as it is shown by Libet (2004) experiments on consciousness and will. In conclusion, the discovery of MNs, as well as giving an immediate biological foundation to the concept of empathy and related ideas, makes us understand, in my opinion, as well observed in the Foreword to Rizzolatti and Sinigaglia (2006), that «The same rigid boundary between perceptual, cognitive and motor processes, ends up being largely artificial: not only perception is immersed in the dynamics of the action, being more articulate and complex than previously thought, but the brain that acts is primarily a brain that understands. This is [...] a pragmatic, pre-conceptual and pre-linguistic, understanding, and yet no less important, since it rests on many of our much celebrated cognitive abilities» (Ibid., p.3).

## References

- [1] Adams, Fred and Ken Aizawa (2010), “Defending the Bounds of Cognition” in Menary 2010, pp. 67-79.
- [2] Chalmers, David (2008), “Foreword” in Clark 2008, pp. ix-xxix.
- [3] Clark, Andy and David Chalmers 1998 “The Extended Mind” *Analysis*, 58, pp. 7-19.
- [4] Clark, Andy (2008), *Supersizing the Mind: Embodiment, Action and Cognitive Extension*. Oxford University Press.
- [5] Clark, Andy (2009), “Letter” *London Review of Books*, 31(6). Retrieved from <http://www.lrb.co.uk/v31/n03/jerry-fodor/where-is-my-mind>.

- [6] Farkas, Katalin (2008), *The Subject's Point of View*. Oxford, Oxford University Press 2008.
- [7] Fodor, J. (2009), Where is my mind? [Review of the book *Supersizing the Mind: Embodiment, Action and Cognitive Extension*.] *London Review of Books*, 31(3), 13-15.
- [8] Gertler, Brie (2007), "Overextending the Mind" in Brie Gertler and Lawrence Shapiro (eds.) *Arguing about the Mind*. New York and London: Routledge, pp.196-205.
- [9] Hurley, Susan (2010), "The Varieties of Externalism" in Menary 2010a, pp. 101-53.
- [10] Menary, Richard (ed.) 2010a. *The Extended Mind*. A Bradford Book Cambridge, Massachusetts: The MIT Press.
- [11] Menary, Richard (2010). "Introduction. The extended mind in focus" in: Menary 2010a, pp. 1-25.
- [12] Pullman, Philip, (1995-2000), *His Dark Materials*. Scholastic
- [13] Sprevak, Mark (2009), "Extended Cognition and Functionalism" *The Journal of Philosophy* (2009), 106, pp. 503-527.
- [14] Wheeler, Michael (2010), "In Defense of Extended Functionalism" in Menary 2010a, pp. 245-69.
- [15] Darwin C. (1872), *The expression of the Emotions in Man and Animals*, Published by John Murray, London.
- [16] Davidson D. (1970), "Mental Events", in Foster, L. e Swanson, J. W. ed., *Experience and Theory*, University of Massachusetts Press, Amhest (1970), pp. 79-101.
- [17] Kim J. (1993), *Supervenience and Mind*, Cambridge University Press, Cambridge.
- [18] Kim J. (1996), *Philosophy of Mind*, Boulder, CO, Westview Press.
- [19] Kim J. (1998), *Mind in a Physical World. An Essay on the Mind-Body Problem and Mental Causation*, MIT Press.
- [20] Kim J. (2005), *Physicalism, or Something near enough*, Princeton University Press, Princeton (NJ).
- [21] Libet B. (2004), *Mind Time. The Temporal Factor in Consciousness*, by the President and Fellows of Harvard College (trad. it.: *Mind Time. Il fattore temporale nella coscienza*, Raffaello Cortina Publisher, Milan 2007).
- [22] Rizzolatti G., Sinigaglia, C. (2006) *So quel che fai. Il cervello che agisce e i neuroni specchio*, Milano: Raffaello Cortina Editore, 2006, translated in *Mirrors in the Brain. How Our Minds Share Actions and Emotions*, Oxford University Press, 2008.
- [23] Simonetti N. (2003), " 'Levels' and 'Orders': The Multi-Layered Metaphysical Model of J. Kim", *Book of Abstracts*, "European Society for Philosophy and Psychology" (ESPP), Tourin, 9-12 July 2003, Fondazione Rosselli.
- [24] Simonetti, N., Zanardi R. (2004), *Filosofia e scienze della mente*, Armando Publisher, Rome 2004.
- [25] Simonetti N. (2009), "Filosofia della mente e neuroscienze: il caso dei neuroni specchio". Text available on: <http://www.phenomenologylab.eu/index.php/2009/11/filosofia-della-mente-e-neuroscienze-il-caso-dei-neuroni-specchio/>.
- [26] Simonetti N. (2010) "Neuroscience and Philosophy of Mind: a Reductive Interpretation of the 'Mirror System' ", in *Practices of Cognition*, edited by Marco Cruciani, University of Trento.
- [27] Simonetti N. (2012), "La mente incorporata. La lezione di J. Kim sino ai neuroni specchio", Aracne Publisher, Rome 2012.
- [28] Simonetti N. (2012), *Supervenience, Reductionism and Mirror Neurons System*, Lambert Academic Publishing (LAP), Saarbrücken, 2012.
- [29] Simonetti, N. (2013) "I neuroni specchio tra neuroscienze e filosofia della mente", 29.09.2013, in *Rivista Elettronica Filosofia e nuovi sentieri*: <http://filosofiaenuovisentieri.it/2013/09/29/neuroni-specchio/>.