

Original article**□ Mirror neurons and free will**

A. LAVAZZA, L. SAMMICHELI*

*Centro Universitario Internazionale, Città di Monte S. Savino (Arezzo), Italy*** Scuola di Psicologia e Scienze della Formazione, University of Bologna, Italy*

SUMMARY: *As among the prerequisites of free will is the intentionality of action (the action must be voluntarily directed towards a specific purpose), the object of analyses of the mirror neuron system, we intend to explore the as-yet speculative difference between free will in the “ontological” sense (“in the third person”, one could say) and free will deriving from a broader “phenomenological” perspective (“in the first person”) of the perception of intentionality in human actions. The former can be traced back to Libet’s famous research on the timing of volitional acts. The counterintuitive - revolutionary according to many - results of Libet’s experiments lie in the comparison of subjective decision timing with neural timing: the brain activity involved in the initiation of action began in the prefrontal motor cortex well ahead of the moment in which the subjects seemed to have made a decision. The “phenomenological” perspective has to do with the psychological “construct” of the intentionality of others. A robust body of literature now exists that tries to comprehend moral reasoning by exploring its potential “pre-moral” constituents, or the construct of the intentionality of others. The idea, tackled from several different perspectives, is that in order to understand moral judgment, one must first understand the mechanisms used to build the “grammar of action”, or the deciphering of observed behaviour. In this sense, the mirror neurons make it possible to understand the actions of others “from the inside”, encoding them in terms of one’s own motor possibilities. In the judicial category of the subjective element of crime, we believe we can identify phenomenological free will, or the phenomenon on the basis of which the actions of others appear to us as free and agential: this is a necessary premise so that these free actions can be considered, upon further mental evaluation, punishable.*

KEY WORDS: *Free will, Intentionality, Mirror neurons, Neurolaw, Phenomenology.*

□ MIRROR NEURONS AND TWO VIEWS OF FREE WILL

“The sight of agony has no effect on me whatsoever. I kill a man as I drink a glass of wine”: this is how the *poète-assassin* Lacenaire summed up his complete lack of empathy. Cesare Lombroso, who studied Lacenaire (along with other famous criminals of the time) while writing his most important book, *L’Uomo delinquente* (Delinquent Man), moves from his case to a blanket statement: what seems to characterize

subjects such as Lacenaire is the selective loss of the ability to empathize with the suffering of others: “The first to disappear is that feeling of compassion for the misfortune of others, which, according to some psychologists, is so deeply rooted in our being”(page 127)⁽¹²⁾. In the emerging field of research known as neurolaw, empathy, and more generally the mechanisms underlying the ability to “mentalize” the behaviour of others, have become of fundamental importance.

Firstly, studies on empathy and the ability to mentalize

are part of a broader, classical field of study encompassing forensic medicine and psychopathology and regarding the analysis of the basic psychic abilities required for fitness to stand trial; in other words, the study of mental pathologies that can qualify a person who commits a crime as legally insane. In particular, the ability to empathize has played a central role in the longstanding debate on the medical and legal status of the nosological category of “psychopathology”. In this regard, the discovery of so-called mirror neurons in the macaque ventral premotor cortex has been extremely important. The mirror neurons discharge during goal-oriented actions, but also when the macaque observes somebody else, for example, grasping an object with a precision grip. The whole mirror mechanism “appears to play a fundamental role in both action understanding and imitation”⁽¹⁴⁾. According to numerous studies, this seems to hold true in humans as well⁽¹⁶⁾.

As “the mirror-neuron system provides the essential physical other-to-self mapping that is necessary for comprehending physical actions of intentional agents”⁽²²⁾, it may be possible to attempt to draw from the phenomenological perception of intent a distinction within the general concept of free will, regarding in particular its application to jurisprudence and neuropsychology. Free will is generally characterized by two conditions: a) the possibility of doing otherwise; in other words, the agent can choose between alternative courses of action; and b) self-determination: the agent determines which course of action to take.

Among the prerequisites of free will is the intentionality of action; in other words, the action must be voluntarily directed towards a specific purpose, which is the object of the analyses of the mirror neuron system. In this paper, we intend to further explore the as-yet speculative difference between free will in the “ontological” sense (“in the third person”, one could say) and free will deriving from a broader “phenomenological” perspective (“in the first person”) of the perception of intentionality in human actions.

The link between the two different approaches to the conceptualization of human free will takes on a certain importance from the privileged point of view of forensic neuropathology. By ‘privileged point of view’, we mean that the practical needs (and socially sensitive function) of judicial activities sometimes seem to be particularly effective in heuristic terms, and bring to the fore certain theoretical questions that might otherwise remain unexplored or overlooked by academic researchers. The point of view of forensic

neuropathology, which necessarily requires pigeonholing in one judicial category or another, seems somehow bound to emphasize the distinction between matters that concern what we may define as “ontological” free will (the problem of the existence of the ability to act freely in humans, or lack thereof) and those which concern what we may define as “phenomenological free will” (the study of the characteristics on the basis of which our actions, and especially those of others, appear to us to be free).

□ “ONTOLOGICAL” FREE WILL

The concept of “ontological” free will can be traced back to Libet’s famous research^(9,10) on the timing of volitional acts. The counterintuitive - revolutionary according to many - results of Libet’s experiments lie in the comparison of subjective decision timing and neural timing: the brain activity involved in the initiation of action is begun in the prefrontal motor cortex well ahead of the moment in which the subject seems to have made a decision. Indeed, volunteers became conscious of their decision to act about 350 milliseconds *after* the onset of type II motor readiness potential (typical of unplanned, spontaneous acts) and 500 to 800 milliseconds *after* the onset of type I motor readiness potential (typical of planned, conscious acts). The volitional process thus seems to begin unconsciously, since the brain readies itself for action much earlier of the moment in which the subject becomes aware of their decision to act.

The results of Libet’s experiments, which had long been questioned but have been confirmed in essence by a series of experiments repeated under stringently controlled lab conditions, seem to indicate - indeed, according to Libet and a great many contemporary scientists, they *prove* - that our actions (or at least the types of actions subject to this research) are caused by preconscious cerebral activity, which is only subsequently registered consciously, on average only 206 milliseconds before the act is performed with regards to intention (the so-called W judgement, after will), and 86 milliseconds before with regards to the act itself (the so-called M judgement, after movement). Many have inferred from these findings that conscious intentions are not behind our volitional actions, because they arrive after cerebral motor preparation activity - which is inaccessible to our consciousness for a certain period of time - and appear only once the process that results in movement has already been

launched. In this light, according to many Authors, Libet's experiments show that intentions are not the true cause of our actions, because the true causal work is performed by neural processes that precede it in time. If this is true, then intentions are causally ineffective or, in technical terms, "epiphenomenal". Without getting into the debate that followed, or discussing the sophisticated research on the subject⁽²¹⁾, here we will merely point out here that in the legal field, Libet's research has often been associated with the construction of the model of the mind in legal subjects: in other words, they have been taken as paradigmatic examples of *new sciences*, which, if taken seriously, could (and should) call into question one of the underlying assumptions of the main Western judicial models: that of the implicit consideration of a penal subject as being free and fit to stand trial (except in the presence of pathological causes) and thus fit to be punished.

In our opinion, this perspective encompasses all the positions - defined as "radical" by others^(2,6,7,18) - of those for whom the possibility of "scientific proof" of the absence of free will necessitate revision of the retribution framework of penal systems. Without burdening our argument with excessive legal references, we will only point out that the legal counterpart of the "ontological" perspective of free will is the concept of imputability (sanctioned by art. 85 in the Italian penal code), whose formal axiom is that the subject liable for penal action must be *imputable*, and thus "in full possession of his or her faculties," i.e., endowed with free will. Imputability as a legal concept thus has to do with problems related to *ontological* free will: if the latter exists (whether proven scientifically or as a legal convention) the subject is imputable; if it is lacking (again, whether according to science or legal convention) the subject is not imputable.

□ **"PHENOMENOLOGICAL" FREE WILL:
THE PERCEPTION
OF INTENTIONALITY**

The other perspective with which *neurolaw* grapples with "freedom" of action in humans has to do with the psychological "construct" of the intentionality of others. A robust body of literature now exists that tries to comprehend moral reasoning by exploring its potential "pre-moral" constituents, or the construct of the intentionality of others.

The idea, tackled from several different perspectives, is that in order to understand moral judgment, one must first understand the mechanisms used to build the "grammar of action", or to decipher observed behaviour. In this sense, the mirror neurons make it possible to understand the actions of others "from the inside", encoding them in terms of one's own motor possibilities. Indeed, "although there are several mechanisms through which one can understand the behaviour of other individuals, the parieto-frontal mechanism is the only one that allows an individual to understand the action of others 'from the inside' and gives the observer a first-hand grasp of the motor goals and intentions of other individuals"⁽¹⁷⁾. More specifically, as stated by Sinigaglia and Rizzolatti, "it seems almost obvious to assume, at least at first glance, that the attribution of actions to the self or to the other should be based on separate neural representations. Two distinct neural networks should underlie our and others' actions. However, it is just this kind of assumption that the discovery of the mirror mechanism has radically undermined. Indeed, what the functional properties of the mirror mechanism tell us is that the self and the other are so strictly intertwined that, even at the basic level, self- and other-attribution processes are mutually related to each other, being both intimately rooted in a common motor ground^(4-5,15,19). More precisely, the mirror mechanism clearly indicates that (i) in order to be attributed either to the self or to the other, actions should be represented as actual motor possibilities for the agent and (ii) the distinction between self and other should stem from their shared motor goals and motor intentions, because it is on the basis of this common motor ground^(4-5,15,19) that we are able to differentiate ourselves from the other selves." (page 69)⁽²⁰⁾.

Based on current knowledge, what emerges is that the grammar of intention is partially grounded in immediate understanding, made possible by the mirror neuron system.

"The chain organization of the cortical motor system provides the mirror mechanism with the possibility to encode not only single motor goals *per se* (e.g., reaching, grasping, holding, etc.), but also motor goals as being intentionally related one to another, thus representing the motor intention with which they might be achieved (e.g. reaching for grasping for bringing-to-the-mouth or reaching for grasping for moving-away). The richness of our motor repertoire does not depend only on the fineness-of-grain of motor goal representation; rather, it essentially relies

on our capability to represent from the inside more and more complex goal architectures, recruiting them both when we perform a given action and when we observe someone else performing it. This capability critically contributes to shaping our experience of ourselves and of other selves, providing us with a multilayered motor representation both of our own and of others' action possibilities." (page 71)⁽²⁰⁾.

Of course, not everything is due to the mirror system, as many studies have shown. For example, "recent research suggests that the inference of others' intentions from their observed actions is supported by two neural systems that perform complementary roles. The human putative mirror neuron system is thought to support automatic motor simulations of observed actions, with increased activity for previously experienced actions, whereas the mentalizing system provides reflective, non-intuitive reasoning of others' perspectives, particularly in the absence of prior experience"⁽¹¹⁾. So, in general, observing all gestures compared to observing still images is associated with increased activity in key regions of both the mirror neurons system and mentalizing systems.

It is also interesting that: "The richer and more diversified our motor repertoire, the sharper our sensitivity to others' actions, so that our capability to make sense of others turns out to be rooted in our capability to make sense of ourselves. It follows that, if more individuals share the same motor repertoire, the richer and more diversified such a motor repertoire is, the more these individuals will be able to be mutually reflected by their own motor possibilities, thus coming to a more and more fine-grained understanding from the inside of each other. In other words, the more individuals share their own motor repertoire with each other, the more fine-grained is the experience they make of action possibilities when these action possibilities are relative both to their own selves and to other selves." (page 71)⁽²⁰⁾.

From a more general perspective, in an analysis of the various dynamics through which the attribution of causality affects the interpretation of actions and omissions, Cushman and Greene⁽¹⁾ (who in other instances adopt positions discussed in the first paragraph), conclude that "the moral distinction depends on processing features of non-moral cognitive processes such as causal attribution and intentional attribution".

Turning to research in the naturalistic-neuroscientific field, a fruitful perspective - which, as we shall see, may have a connection with simulation theory - is

that of *moral grammar*, as proposed by Hauser⁽⁸⁾ on the basis of the ideas Rawls⁽¹³⁾ set out in "A Theory of Justice". This perspective aims to address, through an innatist and biological approach, the problem of the intercultural similarity of moral intuition, and of basic judicial rules regarding murder, incest, theft, etc. In particular, Hauser - albeit with a great deal of emphasis on the evolutionary element (and thus formally far removed from phenomenological tradition; the relationship between phenomenology and naturalism is a highly complex area of philosophical thought, which we cannot tackle here) - touches upon a central issue in the scientific exploration of moral thought, namely the study of the semantic aspect of behaviour as the fundamental premise for any further analysis. The starting point of this approach is the introduction of a structural analogy between the forms of language organization and those of intentional conduct. Indeed, much like language uses discrete elements combined and re-combined to create an infinite variety of expressions carrying meaning (from syllables to words to full sentences), so do "Actions appear to live in a parallel hierarchical universe. Like phonemes, many actions lack meaning. When combined, actions are often meaningful. [...] When actions are combined, they can represent an agent's goals, his means, and the consequences of his action or the omission of an action." (page 47)⁽⁸⁾.

Human behaviour can thus be represented as a text whose structure immediately expresses a "sign-meaning" relationship: if in a text the main relationship is that between the expressive plane (sign) and the content plane (meaning), in intentional behaviour the main relationship is that between manifest behaviour (sign) and its underlying mental states (meaning). The conduct manifested thus becomes a "sign" in terms of "something that stands for something else" (this was the definition of "sign" given in Medieval schools: *aliquid stat pro aliquo*), in this case the underlying mental states. Upon this Hauser bases his argument, which is directly tied to the study of moral psychology. Recalling the idea of universal (and innate) grammar proposed by Noam Chomsky, he suggests that similar principles may also underlie moral behaviour. Much as the existence of a universal grammar relating to the automatic, innate (and universal, in light of the use of constant, inter-cultural principles) nature of language is postulated, so is the existence of a hypothetical moral grammar proposed, with the same set of basic principles. "For language, we recombine words and higher-order combinations

of these words (noun and verb phrases). For morality, we recombine actions, their causes and consequences. Like Chomsky, Rawls suggested that we may have to invent an entirely new set of concepts and operations to describe the universal moral principles.” (pages 47-48)⁽⁶⁾.

Universal moral grammar makes it possible to identify fundamental moral rules underlying the different expressions of cultural systems: the psychological mechanism of assessing a given action is generally shared, albeit with potentially different parameters. And the premise of such a moral grammar must inevitably be likened to a sort of behavioural grammar, a syntax and set of semantic rules through which we interpret the behaviour of others. It is interesting to note how this “semantic consubstantiality” of social behaviour crops up again within a modern phenomenological approach, in its proposal of a “third way” as an alternative to the traditional bi-partition of theories on social cognition⁽⁹⁾. In summary, the two main perspectives are the “theory of theory” and the “theory of simulation”. The former holds that our understanding of others rests upon the psychology of common sense: the behaviour of others, which is unknowable and, in itself, inferred, just like any other natural phenomenon, through the use of a cognitive theory, and, more specifically, this theory sees others as agents endowed with thoughts and intentions. In contrast, simulation theory holds that the understanding of the behaviour of others is mediated by an act of simulation, through which observers put themselves in the shoes of the person being observed: by projecting one’s own mental state, the mental states of others can be understood. An explicit theory is therefore not necessary, since understanding derives directly from the act of identifying with the other.

On the basis of this interpretation of social cognition, the crucial mechanism is neither the use of “theories” of the mind, nor the implementation of “simulation processes”, but rather the immediate carrying out of an intentional behaviour associated with a meaning (that is, endowed with a semantic nature).

“Mirror activation, on this interpretation, is not the initiation of simulation, it’s a part of a direct intersubjective perception of what the other is doing. At the phenomenological level, when I see the other’s action or gesture, I see (*I directly perceive*) the meaning in the action or gesture.” (page 179)⁽⁶⁾. And furthermore: “Expression is more than simply a bridge that closes the gap between inner mental states and external bodily behaviour. While seeing the actions

and expressive movements of the others persons, one already sees their meaning. No inference to a hidden set of mental states is necessary. Expressive behaviour is saturated with the meaning of the mind; it reveals the mind to us.” (page 185)⁽³⁾.

On the basis of this hypothesis, then, the close parallels between the mechanisms of language and those of social cognition lie in the fact that in both cases, the phenomenon encountered by the subject is a *semantic phenomenon*, in which the strictly perspective plane (the ink markings in a written text, the body movements of the other) is augmented by a sign plane, which is associated with something else (the meaning of the written words, the intention of the other). For our purpose, the point that needs to be stressed is that the issue of moral and legal responsibility (and that of free will) finds fertile ground in the microscopic examination of the psychological mechanisms through which we “construct social phenomena”: the way in which our brain leads us to perceive the action of another as determined/non-agential (an involuntary jostle that causes us to trip on a bus), and at another time leads us to perceive it instead as voluntary-agential (a punch thrown by an over-excited fan during a game). The deciphering of this behavioural text precedes any moral judgment (involuntary push = excusable action: “Don’t worry, the driver hit the brakes too suddenly”; vs. a voluntary punch = reproachable action: “I’m going to call the police”).

A clear example of how neuroscience can help elucidate psychological mechanisms can be found in the case of mirror neurons. As argued by Gallagher and Zahavi, they seem to put into action a psychological mechanism that one can readily find (with a minimum of self-examination) in one’s own immediate experiences: we think we are able to understand others by mirroring ourselves in them, or putting ourselves in their shoes. It is only after this very quick “reflex” that we are able to judge - in the broadest sense - the behaviour of others.

In this sense, neurolaw helps us distinguish the various perspectives of the conceptualization of free will. Indeed, if we wanted to limit the above discussion to the legal realm, we realize that the concepts that can be summed up by the idea of phenomenological free will belong to different legal categories from the concepts related to ontological free will. In other words, the ways in which the law deals with the “perception of intentionality” in a crime suspect are those that fall in the category that can be summarized as “culpability”, or a subjective element of the crime.

The law requires that objectively illicit behaviour (*i.e.*, *contra ius*) be accompanied by a morally reproachable mental attitude: in the Italian Penal Code, art. 43 summarizes the three forms of “reproachability”: *malice, guilt and premeditation*. In the Anglo-Saxon legal system, the corresponding legal category for the psychological element involved in crime can be found in the concept of *mens rea*, which is considered an essential element of illicit behaviour, together with the “material” component of *actus reus*. What is of most interest to us in our neurolegal analysis is the fact that this legal category (different from the category of liability discussed with regards to “free will in the third person”) deals with the ways in which law has regulated the evaluation of attribution (and thus also the “construction of intentionality”) with regards to illicit acts. In other words, the law, in an entirely implicit way, independently from any *theory* on moral judgment, has established that the premise for attributing responsibility lies in the fact that the imputable act has been “perceived” by the judiciary body as *intentional* (that is, with a purpose) and that this *intentionality* is reproachable on the basis of the psychological attribution attached to it (*malicious or culpable*).

Thus, in the judicial category of the subjective element of crime, we believe we can identify *phenomenological free will*, or the *phenomenon* on the basis of which the actions of others appear to us as *free and agential*: as we have said, this is a necessary premise so that these free *actions* can be considered, upon further mental evaluation, punishable (that is, deserving of reproach).

□ REFERENCES

- Cushman FA, Greene JD. The philosopher in the theater. In: M Mikulincer, PR Shaver (editors): *The Social psychology of morality: The origins of good and evil*. American Psychological Association, Washington (USA), 2012: 33-50.
- Eagleman D. *Incognito: the secret lives of the brain*. Pantheon, New York (USA), 2011.
- Gallagher S, Zahavi D. *The phenomenological mind: an introduction to philosophy of mind and cognitive science*. Routledge, London (United Kingdom), 2008.
- Gallese V, Keysers C, Rizzolatti G. A unifying view of the basis of social cognition. *Trends Cogn Sci* 2004; 8 (9): 396-403.
- Gallese V, Rochat M, Cossu G, Sinigaglia C. Motor cognition and its role in the phylogeny and ontogeny of action understanding. *Dev Psychol* 2009; 45 (1): 103-113.
- Greene J., Cohen J. For the law, neuroscience changes nothing and everything. In: S Zeki, OR Goodenough (editors): *Law and the brain*. Oxford University Press, Oxford (United Kingdom), 2004: 207-226.
- Harris S. *Free will*. Free Press, New York (USA), 2012.
- Hauser M. *Moral minds: how nature designed a universal sense of right and wrong*. Harper Collins, New York (USA), 2006.
- Libet B. *Mind time: the temporal factor in consciousness*. Harvard University Press, Cambridge (USA), 2004.
- Libet B, Gleason CA, Wright EW, Pearl DK. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain* 1983; 106 (Pt 3) 623-642.
- Liew SL, Han S, Aziz-Zadeh L. Familiarity modulates mirror neuron and mentalizing regions during intention understanding. *Hum Brain Mapp* 2011; 32 (11): 1986-1997.
- Lombroso C. *L'uomo delinquente studiato in rapporto all'antropologia, alla medicina legale ed alle discipline carcerarie*. Il Mulino, Bologna (Italy), 2012.
- Rawls J. *A theory of justice*. Harvard University Press, Cambridge (USA), 1971.
- Rizzolatti G, Craighero L. The mirror-neuron system. *Annu Rev Neurosci*. 2004; 27:169-192.
- Rizzolatti G, Fogassi L, Gallese V. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci* 2001; 2 (9): 661-670.
- Rizzolatti G, Sinigaglia C. *Mirrors in the brain: how our minds share actions and emotions*. Oxford University Press, Oxford (United Kingdom), 2008.
- Rizzolatti G, Sinigaglia C. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat Rev Neurosci* 2010; 11 (4): 264-274.
- Sapolsky R. The frontal cortex and the criminal justice system. In: S Zeki, OR Goodenough (editors): *Law and the brain*. Oxford University Press, Oxford (United Kingdom), 2004: 227-243.
- Sinigaglia C. Mirror in action. *J Consciousness Stud* 2009; 16: 309-334.
- Sinigaglia C., Rizzolatti G. Through the looking glass: Self and others. *Conscious Cogn* 2011; 20: 64-74.
- Soon CS, Brass M., Heinze HJ, Haynes J-D. Unconscious determinants of free decisions in the human brain. *Nat Neurosci* 2008; 11: 543-545.
- Uddin LQ, Iacoboni M., Lange C., Keena JP. The self and social cognition: The role of cortical midline structures and mirror neurons. *Trends Cogn Sci* 2007; 11: 153-157.

DISCLOSURE. *The Authors declare no conflicts of interest.*