Free Will, Neurosciences & Robotics

Sara Fernandes, *Leonor Almeida* and *Alexandre Castro Caldas*

1. *The Neuroethical Problem of Human Freedom*

The neuroscience of ethics is the domain of neuroethics that investigates the problem of human freedom by studying the brain of the brain. Both philosophy and neurosciences strive to understand the biological and mental distinctions between behavior driven by freedom of will and behavior that lacks it. However, it is essential to reflect on the value attributed to neuroscientific discoveries in understanding human beings, particularly concerning the phenomenon of agency in the world. Are the findings of neuroscientific research on human behavior sufficient to offer a philosophical answer to the problem of human freedom and determinism? [1]

Recent advances in neurosciences have enabled a more rigorous systematization of the brain areas involved in reasoning and ethical decision-making. For example, the amygdala detects, evaluates, and assigns emotional significance to an individual's options, while the hippocampus complements this emotional evaluation with autobiographical memory. The anterior cingulate cortex helps us anticipate and solve practical problems, recognize mistakes, and manage situations of uncertainty. Meanwhile, the hypothalamus regulates the body's internal stability and links it to survival behaviors. Finally, mirror neurons allow us to internally simulate actions performed by others or ourselves without carrying them out. As a result, mirror neurons activate whenever an intention is present in an individual's mental process or when they per-

[1] A. Lavazza, *Free Will and Neuroscience: From Explaining Freedom Away to New Ways of Operationalizing and Measuring It*, In *Frontiers in Human Neuroscience*, X, 262, 2016, pp. 1-17.

form intentional behavior [2]. Thus, we understand that the orbitofrontal region is mainly responsible for the social brain. When it is damaged, such as by trauma, the individual may no longer exhibit appropriate behavior (behavior that is socially expected of oneself and others).

We need to revisit an important question: How should we interpret these neuroscientific findings from a philosophical standpoint? Do advances in our understanding of neurobiology undermine or even eliminate the concepts of personal freedom and responsibility? Do these findings imply that humans—whether they are healthy or experiencing a central nervous system disorder—are essentially prisoners of their brains, with their decisions completely determined by this organ, thus invalidating the idea of free will? In essence, to what degree do the findings from neuroscience about human behavior and decision-making challenge the widely held belief in freedom? Or, do they simply help us understand the neurobiological processes that underpin human actions and the concept of freedom? [3]

Rizzolatti's team's unexpected discovery of *mirror neurons* represent one of the most significant breakthroughs in neurosciences in recent decades. This finding has provided significant insights into how the social brain functions, including social skills, learning processes, and the emotional and cognitive aspects of empathy. While studying neurons in a region of the Rhesus monkey brain that controls the hand muscles, the team expected to find neurons that activated when the monkey engaged in specific actions, such as catching a ball or reaching for a banana. They discovered that certain neurons did activate during these actions. However, a surprising observation occurred when the researchers had lunch in the same room as the monkeys. They noticed that some of these neurons also fired when the monkeys watched an experimenter perform the same action, like bringing food to their mouth.

In summary, mirror neurons activate when a monkey either performs an action or observes another—be it a person or another monkey—doing the same action.

This discovery scientifically justifies the importance of intersubjectivity for the constitution of the person and human sociability in general, as Greco-Roman civilization had emphasized on a philosophical

[2] L. Tancredi, *Hardwired Behavior. What Neuroscience reveals about morality*, Cambridge, Cambridge, Cambridge University Press, 2005.

[3] A. Lavazza - S. Inglese, *Operationalizing and Measuring (a kind of) Free Will (and Responsibility). Towards a New Framework for Psychology, Ethics, and Law*, In *Rivista Internazionale di Filosofia e* Psicologia, VI, 1, 2015, 37-39.

ISSN 0039-4130

level for considerable centuries in the West. Mirror neurons, activated whenever someone acts, also enable the internal simulation of behavior practiced by others or oneself without carrying it out. Simply remembering an action or imagining that you or someone else will act can activate these neurons, underscoring their crucial role in understanding human behavior. Mirror neurons are activated whenever there is a mental intention, whether intentional behavior is performed or observed in someone else. Their activation in real and purely imagined scenarios (representing an "as if" experience in others or oneself) suggests they play a significant role in fundamental human traits. These include feeling, sharing, and recognizing emotions in others—traits closely tied to empathy. The ability to empathize, which involves identifying and sharing in others' sadness and joy, is essential for forming distinctly human relationships like friendship and love. It also underpins various forms of learning, both cognitive and social, as imitation serves as a core mechanism in the learning process. [4] As A.Lavazza argues:

> «Until a few years ago, empathy was mainly an object of philosophical and psychological research; then the discovery of mirror neurons, considered by many (though certainly not all) to be a key mechanism in empathy, brought research in cognitive neuroscience to the fore. Having identified the circumscribed brain areas that are activated both when we perform an action and when we observe someone performing that action marked a turning point in the debate on the genesis of understanding and identification with the experiences of others, one of the keys to social life. That empathy can be embodied and primary, almost an automatism that we are all equipped with (unless one has neurological deficits), has challenged many assumptions about the role of education and culture» [5].

2. *The Neuroscientific Research of B.Libet and P.Haggard*

B. Libet and P. Haggard carried out the first neuroscience studies that significantly challenged the belief in human freedom. The first work, published by neuroscientist B. Libet, sought to show that the individual is only aware of his intention to act after the brain has prepared its body for action (potential readiness) [6]. The research showed that the

[4] L. Cattaneo - G.Rizzolatti, *The mirror neuron system,* in *Archives of Neurology,* LXVI, 5, 2008, p. 557-560.

[5] A. Lavazza - S. Inglese, *Operationalizing and Measuring*, cit.

[6] B. Libet, *Unconscious cerebral initiative and the role of conscious will in voluntary action*,

brain prepares the individual for action by activating the motor cortex before they know their intention to act. So, they proved the existence of temporal unconscious brain processes before conscious processes in individuals. Secondly, the study showed that unconscious brain processes set the stage for conscious processes related to human intention, indicating that intentional states cannot exist without these underlying unconscious neurological mechanisms.

In a later study, Haggard and Libet developed this research with more sophisticated measuring instruments, such as the electroencephalogram. They concluded that, in preparation for the action, before the individual became aware of their intention, the brain prepared their body in general and the specific side of the body with which they were going to act, activating the premotor cortex [7]. Thus, in addition to the readiness potential, there was also what the authors called a lateralized readiness potential.

With these studies, the authors brought the debate on human freedom into the scientific field. To put it another way, they used the contributions of neurosciences and empirical psychology to broaden its scope to new research areas. They hope that, since philosophers haven't 'solved' the problem so far, these new areas can do so [8]. Philosophical research is fundamentally conceptual. So, when the two fields intersect, our first task is to understand and clarify—philosophically—what neuroscientists mean by 'being free' when they use this term. In our view, their understanding of being free implies that the entire chain of events leading to an action is under the conscious control of the person performing it. In this context, consciousness is considered a crucial part of free behavior. This means that for an action to be considered free, consciousness must directly cause it.

Libet and Haggard interpreted their findings in a way that challenged the belief in human freedom. They argued that human behaviors, even those that seem free based on the individual's feeling of freedom (connected to the intention and decisions made by the "agent"), are actually influenced by prior events that the individual cannot consciously control. Thus, in the authors' view, these results allegedly show that we do not consciously cause our intentions, decisions, and volitions and, in this sense, our actions, so we cannot consider ourselves free.

In *Behavioral and Brain Studies*, VIII, 4, 1985, p. 529.

[7] P. Haggard - B. Libet, *Conscious Intention and Brain Activity*, In *Journal of Consciousness Studies*, VIII, 11, 2001, pp. 47-64.

[8] A. Lavazza - S. Inglese, *Operationalizing and Measuring*, cit., p. 38

ISSN 0039-4130

Libet and Haggard also drew attention to another research result, which has caused immense surprise and reflection in the scientific community ever since. The researchers observed that whenever the intention (to touch the button) became conscious in the experimental subject, they still had a very short period to inhibit their conscious intention and, therefore, not carry out the intended movement (touching the button). The authors suggest two possibilities regarding the belief in an agent's conscious freedom. One possibility is that this belief is an illusion, as conscious actions are influenced by unconscious mental states that can even override the agent's intentions. Alternatively, they propose a more remote hypothesis: the agent may retain some control over their behavior, which is primarily expressed through their initial intention to block certain movements. However, the neural correlates and genesis of this (minimal) kind of self-control (as a veto, denial of intention, and the corresponding movement) have yet to be determined [9]. Based on empirical evidence, those were strong reasons to abandon the widespread, deeply held belief that human beings are free, i.e., possess the ability to initiate their actions through entirely conscious and self-controlled free will.

More recent research, conducted by a team of neuroscientists led by John-Dylan Haynes, has identified the emergence of both behavioral and abstract choices, such as the simple act of raising a hand or performing small mathematical calculations (like addition and subtraction), seconds before the experimental subjects were even aware of their intentions. Furthermore, even though the study involved basic tasks, it suggests a future possibility of being able to "read" our minds using techniques like functional magnetic resonance imaging. This could allow us to know our upcoming choices, thoughts, and mental states—essentially, our private mental experiences—even before we consciously realize them [10].

These neuroscientific studies most significantly challenged the belief in human freedom. They showed that the individual is only aware of his intention to act after the brain has prepared his body for the action (potential readiness). Without these unconscious neurological mechanisms, there could be no intentional mind state. Thus, to neuroscientists, these results show that we do not consciously cause our

[9] A. Lavazza, *Free Will, and Neuroscience*, cit., p. 14.

[10] C.S. Soon - M. Brass - H.J. Heinze - J.D. Haynes, *Unconscious Determinants of Free Decisions in the Human Brain*, in: «Nature Neuroscience», XI, 5, 2008, pp. 543-545.; C.S. Soon, A.H.He, S. Bode, J.D. Haynes, *Predicting Free Choices for Abstract Intentions*, in: «Proceedings of the National Academy of Sciences», CX, 15, 2013, pp. 6217-6222. A. Lavazza - S. Inglese, *Operationalizing and Measuring*, cit., p. 40.

intentions, decisions, volitions, or actions. So, we should not consider ourselves free. From a philosophical point of view, Libet & Haggard defend determinism, challenging a specific conception of freedom, libertarianism. For the followers of the libertarian current, to be free is to be free from any external or internal constraints. Nothing beyond the agent's conscious motivational states can influence his choices and actions without diminishing his capacity for action.

However, it is not clear that the existence of unconscious brain processes, temporally before conscious mental states, implies the conclusion that we are not free or responsible agents. For four reasons:

1) We can interpret the temporal precedence of unconscious mental processes to the consciousness of intention as a preparation of the human organism to form intentions and make decisions. Nor could we expect any other work from the brain.

2) Libet and Haggard's argument seems to be a typical example of the post hoc fallacy, reasoning that invalidly infers a causal relationship between A and B, namely, the conclusion that A is the cause of B, solely because there is a relationship of temporal precedence between event A and event B. Neurological clinical practice can provide supporting arguments for this claim. So-called panic disorder episodes represent an apparent dissociation between unconscious brain mechanisms and conscious mental activity. In such cases, the body reacts as if there were a stimulus capable of triggering fear, even without that stimulus. Conscious activity does not follow this sequence of events, resulting in a dissociation that supports the idea of a lack of causal connection.

3) Given the significant advances in brain sciences over the past decades, with increasingly precise methods and findings, there no longer seems to doubt that the brain is necessary for mental states and most human behavior. Experimental research has made substantial contributions to rethinking the problem of human determinism. However, we must be cautious in how we interpret experimental data. Even if unconscious mental processes always precede conscious mental states, this does not necessarily imply a reduction in freedom. This statement is unequivocal when applied to conscious mental states like emotions. When we feel an emotion directly, we also know that at a later stage, we always have the possibility of counteracting fear and having various possible conscious behaviors instead of being paralyzed, such as running away, avoiding, and cautioning [11].

[11] W. Glannon, *Bioethics and the Brain.*, Oxford University Press, Oxford 2007, pp. 55-56.

ISSN 0039-4130

4) Neurosciences have only empirically discussed free will, i.e., the choice between possible alternatives. It didn't discuss the true freedom that Kant, Paul Ricœur, or Charles Taylor defend, which interests most from an ethical and legal perspective. Under the influence of Kant's philosophy, we believe freedom is fundamentally an experience of self-determination. This notion presupposes a positive conception of freedom, where being free is understood as the will being a cause unto itself. To be free is to align oneself with our actions and to assume their authorship. In this framework, self-determination and self-creation are inseparable.

The most significant decisions in life are not necessarily those that involve choosing between alternatives of equal value. As C.Taylor argues, the capacity to make strong assessments and craft personal life choices is integral to identity.[12] Without this, an agent would lack the depth essential to human nature. The strong evaluator can be profound, as their evaluations are not merely driven by achieving goals but by the life they aim to shape. This project is intimately tied to personal identity, as it depends on a horizon of values that one embraces. Returning to Kantian ethics, the imperative is that every person, in every action, should reflect on whether the maxim of their action could be universal. This reflection affirms our autonomy and dignity, especially when we disobey unethical or unjust demands. We are highly free as we disobey immoral demands.

In light of the current debate about AI and its ethical implications, we emphasize that, based on our perspective on agency and freedom. We believe that traits like intelligence, emotions, awareness, intentionality, practical reasoning, and the ability to make strong and weak evaluations are essential to human experience. Therefore, AI should be regarded as a simulation of these human traits. From both anthropological and ontological standpoints, there are significant differences between real entities and those that are merely simulated. This distinction holds ethically as well: living a genuine life, fully integrated into the world, open to the vulnerability and richness of human relationships—including love, friendship, disappointments, and anguish—is vastly different from a life of simulation. As Nozick's invention of the experience machine suggests, a life of safety, pleasure, and disconnection from real life may be satisfying in some domains. Still, it would lack authenticity because it is artificial.

[12] C. Taylor, *Human Agency and Language. Philosophical Papers I,* Cambridge, Cambridge University Press 1985, pp. 66-68, 73.

ISSN 0039-4130

An example of this dilemma arises in clinical settings, where full-body scans and immediate AI-driven diagnoses are considered potential replacements for human doctors. While AI is an increasingly sophisticated tool for human benefit, it remains an artefact. The clinical relationship, however, is deeply contextual, and AI's guidance does not consider the context of the patient's situation. Human beings develop through their relationships with the world and through communication. Unless an individual prefers to make decisions in complete isolation, they may choose the counsel of a human practitioner. In general, human beings are relational, and in clinical situations, most would prefer a real dialogue and a shared responsibility between patient and doctor.

If a robot is not free in the sense we define it, and thus not an agent, can it be ethical? Alan Winfield programmed a robot to make decisions based on Kantian principles—that all lives are inherently valuable and equal. For Kantians, this approach is ethical but not for consequentialists. However, Winfield challenges the robot with a catastrophic situation where saving everyone is impossible. Based on Kantian ethics, the robot attempts to save everyone but ultimately fails to save anyone. [13] Unlike the programmed robot, this scenario highlights a key human trait: the ability to adapt to new circumstances, which is tied to intelligence—the capacity to solve novel problems by creating new solutions. However, based on fixed principles, a robot is programmed to respond identically each time. It is an ethical zombie. While our character may firmly adhere to certain principles over others, we believe flexibility and practical wisdom - phronesis, as Aristotle and Ricœur suggest - are essential for sound decision-making.

AI is a pale copy of human intelligence. Humans cannot be artificially replicated, and what AI does is merely a simulation—often with extraordinary capabilities, even surpassing our own. While this is true, and despite its limitations, as we have seen earlier, the medical possibilities opened by AI remain fascinating. For example, the basic premise underlying all neuroprosthetic approaches is that targeted and controlled electrical stimulation of nerves or muscles can potentially restore the physiological function of a damaged organ or limb. The continuous development of brain-machine interfaces offers remarka-

[13] A.Winfield - C. Blum, W, Liu, *Towards an Ethical Robot: Internal Models, Consequences and Ethical Action Selection*, In: M. Mistry - A. Leonardis et al. (eds) *Advances in Autonomous Robotics Systems*. TAROS 2014. Lecture Notes in Computer Science, 8717. Springer, Cham. https://doi.org/10.1007/978-3-319-10401-0_8, pp. 9-10.

ISSN 0039-4130

ble hope for patients suffering from a wide range of conditions. Although AI, neurosciences, and robotics cannot provide solutions to every problem or for every person, we have some solid reasons to feel optimistic about their future in healthcare. [14]

In conclusion, neuroscientific research offers new insights into human freedom and determinism philosophical debate. Findings from Libet, Haggard, and Haynes suggest unconscious processes precede conscious intentions but do not eliminate our freedom. Instead, they invite a deeper reflection on how unconscious mechanisms support conscious decision-making. Rooted in Kantian ethics and enriched by P.Ricœur and C.Taylor, true freedom lies in aligning actions with one's values and taking responsibility for them. Similarly, discovering mirror neurons emphasizes intersubjectivity and empathy as essential to ethical behavior.

While AI provides powerful tools for healthcare and problem-solving, it remains a simulation of human intelligence, lacking self-awareness, adaptability, and moral reasoning (ethical zombie). These limitations become evident in relational, context-dependent fields like medicine. Ethical dilemmas posed by AI highlight the inadequacy of rigid, principle-based reasoning, as sustained by Winfield, reaffirming Aristotle's and Ricoeur's emphasis on practical wisdom (*phronesis*). Ultimately, advancements in neurosciences, AI, and robotics must be grounded in a philosophical understanding of freedom, agency, and moral responsibility, preserving the irreplaceable depth of human experience.

References

A. Ferreira - L.Almeida, *Que futuro para a cirurgia oftalmológica?*, in L. Almeida (edited by), *Escrevinhar a Pensar a Bioética. Assuntos de Ética e Direito Médicos*, Loures, Thea Portugal, SA, 2017, pp. 87-88.

A.Winfield, C. Blum, W, Liu, *Towards an Ethical Robot: Internal Models, Consequences and Ethical Action Selection*, in M. Mistry - A. Leonardis et al. (eds), *Advances in Autonomous Robotics Systems*. TAROS 2014. Lecture Notes in Computer Science, 8717. Springer, Cham. https://doi.org/10.1007/978-3-319-10401-0_8.

[14] A. Ferreira, L.Almeida, *Que futuro para a cirurgia oftalmológica?*, in L.Almeida (edited by), *Escrevinhar a Pensar a Bioética. Assuntos de Ética e Direito Médicos*, Loures, Thea Portugal, SA, 2017, pp.87-88.

A. Castro Caldas, *Viagem ao Cérebro e a algumas das suas competências*, Universidade Católica Portuguesa, Lisboa 2008.
A. Lavazza - S. Inglese, *Operationalizing and Measuring (a kind of) Free Will (and Responsibility). Towards a New Framework for Psychology, Ethics, and Law,* in *Rivista Internazionale di Filosofia e* Psicologia, VI, 1, 2015, pp. 37-55.
A. Lavazza, *Free Will and Neuroscience: From Explaining Freedom Away to New Ways of Operationalizing and Measuring It,* in *Frontiers in Human Neuroscience*, X, 262, 2016, pp. 1-17.
Aristotle, *Nicomachean Ethics*, trans. H. Rackham, Cambridge, Harvard University Press.
B. Libet, *Unconscious cerebral iniciative and the role of conscious will in voluntary action*, in *Behavioral and Brain Studies*, VIII, 4, 1985, pp. 529-566.
B. Libet, *Do we have free will?*, in *Journal of Consciousness Studies* VI, 8- 9, 1985, pp. 47-47.
H. Doucet, *Anthropological Challenges Raised By Neuroscience: some ethical reflections,* in *Cambridge Quarterly of Healthcare Ethics*, XVI, 2007, pp. 219-226.
L. Cattaneo - G.Rizzolatti, *The mirror neuron system,* in *Archives of Neurology,* LXVI, 5, pp. 557-560.
L.Tancredi, *Hardwired Behavior. What Neuroscience reveals about morality,* Cambridge University Press, Cambridge 2005.
W. Glannon, *Bioethics and the Brain*, Oxford University Press, Oxford 2007.
W. Glannon, *Free Will and the Brain. Neuroscientifc, Philosophical, and Legal Perspectives*, Cambridge University Press, Cambridge 2015.
P. Haggard - B. Libet, *Conscious Intention and Brain Activity*, in *Journal of Consciousness Studies*, VIII, 11, 2001, pp. 47-64.
P. Churchland, *Neurophilosophy: Toward a Unified Science of the MInd-Brain*, MIT Press, Cambridge and Massachusetts 2002.
Kant, *Critique of Practical Reason*, in *Cambridge Texts in the History of Philosophy*, Cambridge University Press, Cambridge 2015.
P. Ricœur, *Soi-même comme un autre*, Éditions du Seuil, Paris 1990.
C. Taylor, *Human Agency and Language. Philosophical Papers I*, Cambridge University Press, Cambridge 1985.
C. Taylor, *The Ethics of Authenticity*, Harvard University Press, Cambridge 1991.

ISSN 0039-4130