

A NEUROETHICAL APPROACH TO COMPATIBILISTIC FREE WILL

Michela Andrisani

michela.andrisani@edu.unito.it

**1. Introduction 2. Compatibilism: An overview
3. Where, then, can free will be? 4. Conclusion**

ABSTRACT: A NEUROETHICAL APPROACH TO COMPATIBILISTIC FREE WILL

This article reassesses the viability of free will in light of contemporary neuroscience, arguing for a compatibilist model grounded in neurobiological determinism. Rejecting libertarian accounts, it reframes free will as a functional capacity within causal processes, drawing on the identificationist tradition, the reason-responsiveness model, and the Principle of Alternative Possibilities (PAP). Empirical decision-making models, such as conditional intention and integration to bound (COINTOB), together with Jesse Prinz's Attended Intermediate-level Representations theory (AIRs), support the causal role of conscious deliberation. Free will is thus conceived as a scalable capacity embedded in the deterministic structure of cognition.

KEYWORDS: Neuroethics; Free will; Compatibilism; Identificationism; Conscious deliberation.

1. Introduction

Historically, the free will debate has always revolved around the tension between determinism—the view that all events, including human actions, follow from prior causes—and the intuition that individuals can freely choose among alternatives. This tension has generated multiple theories redefining freedom and control within natural laws. The question remains central across

* Ha conseguito la Laurea Magistrale in Filosofia presso l'Università degli Studi di Torino.

philosophy, ethics, psychology, and neuroscience, which now offers empirical insights into the concept of agency.

The first renowned experiment related to free will was conducted by Benjamin Libet and his colleagues in 1983. Participants performed spontaneous movements (such as bending a finger) while reporting the instant of intention. They monitored cerebral activity using EEG, and the outcomes—along with Libet’s unique interpretation—fostered extensive debate on free will, characterized by objections, replications, and revisions. Brain recordings revealed that neural activity, the readiness potential (RP), preceded conscious awareness of intent by several hundred milliseconds¹. Libet concluded that brain processes initiate actions before conscious decision-making, suggesting that free will may be illusory. He admitted only a *conscious veto* over automatic impulses. In essence, it is indeed possible to choose not to engage in the actions that arise in our consciousness, even if they were previously instantiated within our neural system. Though widely debated and methodologically contested, this experiment served as a catalyst for subsequent discussions and fundamentally reshaped how consciousness and volition are addressed. Therefore, it remains foundational for the topic.

Philosophically, the spectrum of positions can be divided into compatibilists, who argue that free will coexists with determinism, and incompatibilists, who maintain that genuine freedom requires indeterminism. The latter group includes libertarians, who defend indeterministic free agency, and hard determinists, denying it entirely. In recent years, findings from neuroscience and cognitive science have shown that mental processes are inseparable from their physical foundation. In other words, if we genuinely wish to comprehend and discuss free will, we must engage with the latest findings in neuroscience. This interdisciplinary reasoning underlies neuroethics—officially born in 2002 in the *Neuroethics*:

¹ Specifically, 550 milliseconds before the movement and 250 milliseconds before their conscious awareness. Cfr. B. Libet, C.A. Gleason, E.W. Wright, and D.K. Pearl, *Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act*, in «Brain: A Journal of Neurology», 106, 1983.

Mapping the Field (NMTF) conference in San Francisco, which featured some of the most distinguished minds in philosophy and sciences².

Following this approach, the present work examines free will through both philosophical and scientific lenses, especially neurobiology and cognitive sciences. In this regard, before moving to the next section, it is essential to establish that, as a methodological premise, I will adopt neurobiological determinism, grounding it in the most up-to-date scientific accounts, specifically drawing on Robert Sapolsky's systematic exposition³. His position—definable as hard incompatibilist—views behavior as fully determined by neural mechanisms shaped by genetic and epigenetic influences that preclude any form of ontological indeterminacy. This is because, although the debate regarding indeterminism is complex and constantly evolving, current libertarian appeals to indeterminism—whether based on chaos theory, emergence theory, or quantum mechanics—do not withstand scrutiny when applied to brain mechanisms and, by extension, to the will. While I share Sapolsky's deterministic premise, I diverge from his conclusions. Determinism, rather than eliminating free will, provides the foundation for a compatibilist model in which psychological capacities—neural in nature—retain explanatory and moral significance. Free will is thus conceived as an internal function within causal processes.

In this context, my approach is also aligned with *neuro-naturalism*⁴. From this perspective, mental states coincide with neural processes governed by natural laws⁵ without interfering with a compatibilistic view of free will or negating their psychological significance. Nahmias's proposal will prove to be

² Cfr. S. Marcus (ed.), *Neuroethics: Mapping The Field. Conference Proceedings*, The Dana Press, New York 2002.

³ More precisely, to what he elaborates in R.M. Sapolsky, *Determined: Life without free will*, Random House, New York 2023.

⁴ As formulated by Eddy Nahmias in E. Nahmias, *Your brain as the source of free will worth wanting: Understanding free will in the age of neuroscience*, in «Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience», 252, 2018.

⁵ Differently from Sapolsky's determinism, this position does not really take a position against indeterministic events and brain.

beneficial for the more detailed discussion on neuroscience and reasoning that will occur later.

The following discussion develops a compatibilist model of free will grounded in determinism, re-examining key conceptual frameworks—particularly the identificationist and reasons-responsiveness models—to establish a coherent link between philosophical theory and neuroscientific evidence. While acknowledging the richness and complexity of the debate, the next paragraph will briefly outline the ontology of compatibilism, as it is particularly relevant to the current discussion.

2. Compatibilism: An overview

The compatibilist view arose, predominantly within British philosophy in the second half of the seventeenth century and developed into the twentieth, when substantial critiques began to surface. As Bernard Berofsky notes, while lacking a formal definition, the cluster of compatibilist theories that emerged around 1700 and flourished until the second half of the twentieth century may be aligned with the term classical compatibilism⁶.

Its central claim—common to its major proponents, including Hobbes, Locke and Hume—is that:

an agent is free if and only if he acted in the way that he did because he wanted (or chose) to do so and would have acted otherwise if he had wanted (chosen) to⁷.

This definition embodies an early form of dispositionalism centred on the absence of external constraints. In this context, one is considered free only when actions follow from one's will, unimpeded by external forces. This aspect, though later refined, remains pivotal in discussions of determinism.

⁶ B. Berofsky, *Classical Compatibilism*, in K. Timpe, M. Griffith, N. Levy, Routledge (eds.), *The Routledge Companion to Free Will*, Routledge, New York 2017. p. 41.

⁷ *Ibid.*, p. 42.

a. *The Principle of Alternate Possibilities*

A second key principle of classical compatibilism is counterfactual power, encapsulated by the Principle of Alternate Possibilities (PAP). The latter posits that moral responsibility, which in this context can be aligned with free will, depends on the possibility of having acted differently. Counterfactual power is the main element of the stronger version of the so-called "forking paths" model⁸, when adopted by compatibilists: free will entails the presence of several actionable "paths" and the power to take the one preferred. Common intuitions, such as compassion for a drug addict or a coerced individual, often reflect this view. As it will be shown, as the century passed, new phenomena have been recognized as blocking alternatives, making the analysis of the problem more and more insidious. Traditionally, PAP has garnered a fairly uniform agreement for its purported self-evidence. For example: if I decide not to go to work, unaware of a protest blocking all traffic⁹, am I free? According to classical compatibilists, such as Locke or Mill, I am not. This is because I lack the possibility of acting otherwise; if, after a brief deliberation, I wanted to go to work instead, I could not be able to do so. Hence, this would violate the aforementioned dispositional condition of freedom. The latter observation underscores an important detail specific to Locke's conception, namely that voluntariness is a necessary but not sufficient condition for free will. The approval of this constraint reached the beginning of the twentieth century, toward the demise of the classical sense of compatibilism, whose most notable figure in this regard is G.E. Moore. However, not all classical compatibilists agreed. To mention one of the most significant opponents to this requisite, Hobbes rejected PAP altogether: for him, as long as an act is voluntary and lacks external constraints, it is free; the origin of the will is irrelevant. To clarify, returning to the previous example, the mere fact that I voluntarily decide to stay at home is the only factor that matters for assessing my action as free; not only that, even if someone manipulated or hypnotized

⁸ *Ibid.*, p. 43.

⁹ This is a revisitation of the fourth example Locke made in J. Locke, *An Essay Concerning Human Understanding*, London 1690.

me, I would remain free. The ability to do otherwise will remain a constant controversy also in the non-classical compatibilist field, as it will be pointed out below.

b. The identificationist solution and the reason-responsiveness account

It is evident that classical compatibilism, while remaining a milestone in the debate, lacks significant conceptual insights. The most critical feature of this paradigm is the underlying account of self-determination. Simply acting on desires without external obstacles does not guarantee freedom; such desires may arise from internal constraints. Today, these latter have been increasingly recognized as inescapable; states such as psychiatric disorders are considered limits to freedom. For example, the desire to spend all one's money in an online poker game could be dictated by an existing precondition of pathological addiction, rather than by a "genuine" disposition to act. Consequently, another issue arises: is the phenomenological manifestation of the mind itself—not solely clinical conditions—constitutionally determined in such a way that it leaves no room for free will?

To address the issue concerning self-determination, Harry Frankfurt reformulated compatibilism in 1971 with his Identificationist view:

[...] one is free to the extent that one is able to do what one truly wants to do, where true wants are understood as desires with which the agent identifies¹⁰.

Here, mere desire is insufficient for freedom; the agent must identify with that desire—what is truly meant by "identify" is not obvious, and it will be specified later. The game addict who bets another hundred dollars may act on a desire, but lacks freedom if he does not endorse that desire. This view comes in broad and narrow forms. The broad view allows PAP to play a role; since agents may be influenced by reasons with which they do not identify, possessing the ability to do otherwise is a crucial element. Being free, thus, means—in the broad identificationist view—acting in accordance with one's

¹⁰ A. Jawroska, *Identificationist Views*, in *The Routledge Companion to Free Will*, cit., p. 15.

identification *or* having the possibility to act driven by one's identification. These two disjunctive meanings can appear quite similar, yet they hold a significant difference: the first does not account for counterfactual power, while the second does. To further explain the first sentence, in cases where the ability to do otherwise lacks, if the subject identifies with the action, he is nevertheless free. Recalling the scenario in which the gambling addict repents the choice of betting, if he, conversely, recognized himself in those actions, he would be free. This is because, although he may not possess the capacity to act differently, the resulting action aligns with his identification in desiring to play another round of poker.

This first proposition, encompassed within the broad view, constitutes the foundation of narrow identificationism. The latter is the version endorsed by Frankfurt, for whom PAP is entirely irrelevant in evaluating an agent's freedom. He distinguishes between acting on one's own free will and the freedom of the will; the former pertains to the alignment of the self with the volitions that drive the action, while the latter coincides with the ability to shape the action according to the motivations with which one identifies. For Frankfurt, possessing the power to act otherwise, i.e. exercising freedom of the will, is not pertinent. He constructs his theorization through a hierarchical model of identification: identifying with a volition that motivates an action entails possessing higher-order desires that align with it. In this way, freedom of the will is not a prerequisite for being considered a free agent; it is both necessary and sufficient to endorse one's own volitions, which harmonize with one's deepest values¹¹.

Frankfurt's model introduces an important distinction between being the origin of an act and endorsing it reflectively; however, it remains vulnerable. In several instances, an agent's higher-order volitions can themselves be shaped by manipulation, coercion, or social conditioning. Does

¹¹ He also asserts that PAP is mistaken in attributing moral responsibility. As outlined above, a person may identify with an action even in circumstances in which she could not do otherwise. For more details on the argumentation, cfr. H. G. Frankfurt, *Alternate Possibilities and Moral Responsibility*, in «The Journal of Philosophy», 66, 1969, pp. 829-839.

their identification still constitute genuine autonomy? As Gary Watson argues, in Frankfurt's account, the agent is equated to second-order desires, yet freedom requires more than internal harmony; it involves a critical self-evaluation of one's motivations¹².

This perspective suggests that an agent must critically assess and endorse their desires based on rational evaluation, rather than solely on higher-order desires; this, in my view, reintroduces PAP as a key element in delineating a plausible account of free agency. The subject appears to demonstrate freedom in a more convincing sense when she is not entirely overwhelmed by other forces, but can instead reveal glimmers of alternative possibilities, introducing evaluation. Several philosophers have developed distinct accounts of identification, conceived as an exercise of deliberation; here, the reason-responsiveness perspective will be considered. By emphasizing an agent's capacity to respond to rational considerations, it provides a more comprehensive framework for understanding free will. While the internal debate is extensive and merits a thorough review, only the essential arguments are presented here.

Following Fischer and Ravizza's Moderate Reasons-Responsiveness (MRR) theory¹³ Michael McKenna describes two key faculties that relate to each other through an asymmetrical relationship¹⁴. Being able to act in accordance with reasons necessitates the possession of both reasons-receptivity and reason-reactivity. The former is the ability to recognize the reasons for acting, while the latter pertains to the ability to react according to

¹² As Michael E. Bratman recalls in M.E. Bratman, *Valuing and the Will*, in «Philosophical Perspectives», 14, 2000, pp. 249–265. Here is also present the distinction with the first Watson's account of "value", coincident with good judgment, and its subsequent revision. Watson recognizes that evaluative process is not necessarily morally connoted, for someone can value and endorse a motivation which does not coincide with the better rational judgment.

¹³ J.M. Fischer, M. Ravizza, *Responsibility and control: A theory of moral responsibility*, Cambridge University Press, Cambridge 1998, pp. 69-76. This compatibilist account adopts the notion of guidance control, according to which moral responsibility depends on actions issuing from the agent's own moderately reasons-responsive mechanism, rather than on the ability to act otherwise under identical conditions.

¹⁴ M. McKenna, *Reason-Responsive Theories of Freedom*, in *The Routledge Companion to Free Will*, cit., pp. 35-36.

the relevant reasons chosen during the deliberation process. The asymmetry lies in the fact that receptivity is primary: agents need not to act on every recognized reason, but must be open to reception. Hence, McKenna states:

[...] moderate reasons-receptivity and a degree of moderate reasons-reactivity that may but need not be weaker than the receptivity component. Regardless, the reactivity component must be moderate enough to display a stable, sane way of relating to a sufficiently robust spectrum of reasons¹⁵.

Through this formulation, McKenna affirms the capacity to respond to sufficiently compelling reasons that support a different course of action, due to a rational evaluative process. Simultaneously, he acknowledges that one can still choose not to act in that manner: if an agent were automatically or mechanically reactive to every recognized reason, their actions would lack the flexibility necessary for freedom. Moreover, reaction could never exceed reception, as it would not be determined by a rational deliberative process; consequently, flowing to an implausible notion of total self-determination. This theorization proves fruitful for the present analysis, as it underscores the importance of deliberation and provides deeper insight into the relationship between behavior and the degree of awareness of reasons. Reasons-reception thus unfolds a set of potential alternative possibilities within a proper process of deliberation¹⁶. This does not imply, however, that the agent is able to choose from alternate possibilities in every case; this would not suit a compatibilist view, nor is it required. Recepting reasons extends the available choices, allowing for better deliberation; without them, possibilities are extremely limited to impulsive decisions.

¹⁵ *Ibidem*.

¹⁶ This is not included in the considerations of McKenna or Fischer and Ravizza, for their account tries to explain why PAP is not necessary for free action. I would like to specify that I agree with the fact that the acceptance of PAP they consider is not necessary for a compatibilist free action. For a further explanation, cfr. D. Robb, *Moral Responsibility and the Principle of Alternative Possibilities*, in E.N. Zalta, U. Nodelman (ed.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, 2023. Here I try to explain a different meaning with which PAP could be intended.

There is one final issue worth assessing: even this view risks depicting the self as a decision-maker separate from its desires. In other words, it fails to account for all the reasons that pertain to the internal world of a subject. Consequently, the previous theories imply a conception of the individual as "divided"¹⁷. Angela Smith refines this further, rejecting the dualism implied in such models¹⁸. She, in fact, states that:

[...] we need a richer conception of moral agency than the one provided by the volitional model. Though the activities of reflective evaluation, endorsement, and rejection highlighted by the volitional model are important expressions of our moral agency, the real core of our moral agency resides in our more basic capacity to evaluate—in complex, spontaneous, and sometimes even contradictory ways—the world around us. This evaluative capacity is manifested in what we unreflectively think, feel, desire, and notice, as well as in what we reflectively choose, endorse, or reject. I will call this alternative model of moral agency the evaluative model¹⁹.

Here, freedom arises not from aligning higher-order desires with action, but from our capacity to evaluate and integrate both reflective and unreflective attitudes. Internal desires are not obstacles to be overcome; instead, they should be acknowledged and understood just as much as external reasons. According to Smith, recognizing and integrating these attitudes enables individuals to deliberate with greater self-awareness, thereby fostering personal growth and moral agency.

This unveils a subtle realization: not only can pathological or externally coercive situations undermine conscious self-determination, but also the ordinary unawareness of one's own desires and attitudes, their origins, and

¹⁷ A criticism that echoes the *Cartesian's theatre* concept described by Daniel Dennett, in which conscious will and unconscious impulses result in a subjective separation that seems implausible. For further details cfr. D.C. Dennett, *Freedom evolves*, Penguin, London 2004, pp. 228-243.

¹⁸ A.M. Smith, *Conflicting Attitudes, Moral Agency, and Conceptions of the Self*, in «Philosophical Topics», 32, 2004, p. 340.

¹⁹ *Ibidem*.

the inability to incorporate them into the deliberative process. This, in turn, emphasizes that we are determined in a more profound sense than we usually think.

Integrating Smith's evaluative model with the reason-responsiveness framework bridges the gap between cognitive deliberation and affective structure, offering a more comprehensive compatibilist account. Integrating Smith's evaluative model with the reason-responsiveness framework bridges the gap between cognitive deliberation and affective structure, offering a more comprehensive compatibilist account²⁰.

In conclusion, it is worth reiterating the key elements that emerged from the analysis of compatibilism. While PAP often remains contested, important compatibilist tools—Frankfurt's distinction between "own free will" and "freedom of the will"²¹, the reasons-responsiveness account, and the integration of affective and rational structures—provide a promising foundation for defending free will within a deterministic framework.

Now, all this considered, comes a crucial point: advancements in neuroscience have created opportunities to scientifically inform reasoning concerning this topic. As De Caro asserts:

To say that the free will discussion must proceed independently from scientific investigations seems in short an untenable thesis[...]. Actually, one thing we can be certain of: that is, with respect to the problem of freedom, the conceptual and empirical planes are firmly intertwined—and whoever forgets it too far can never reach²².

In the following part of the article, I will aim to integrate recent neuroscientific findings with philosophical analysis, to explore a viable approach for further examining the possibility of free will.

²⁰ These statements are analogue to more recent theories about the interplay of emotional and cognitive cortical areas in elaborating moral judgments. Cfr. G. Corbellini e E. Sirgiovanni, *Tutta colpa del cervello: un'introduzione alla neuroetica*, Mondadori, Milano 2013.

²¹ As it will be pointed out, Neil Levy makes a similar distinction when considering in which sense a person should be considered free even when determined by her own mechanisms.

²² M. De Caro, *Libero arbitrio e Neuroscienze*, in A. Lavazza and G. Sartori (eds.), *Neuroetica*, il Mulino, Bologna 2011, p. 102. The english translation is made by me.

3. *Where, then, can free will be?*

In this section, the compatibilist principles discussed above are re-evaluated in light of neuroscientific and neuroethical evidence. Since consciousness plays a central role, it is first necessary to clarify the concept of the self adopted here, ensuring coherence with deterministic premises and avoiding any dualistic interpretations.

a. *There is no self*

The provocative title refers to a specific challenge that neuroscience poses to the classical notion of the self. It undermines the idea of an inner agent directing our actions, known as the *homunculus fallacy*, as described by Neil Levy. The notion of a control center in the brain fails to adequately explain how our minds function²³. The consequence of assuming this concept is an infinite regress when attempting to explain mental functioning.

Even compatibilist theories often assume a self distinct from impulses or first-order desires. Such an account of the self fails to acknowledge that—as previously discussed through Sapolsky’s work—all mental phenomena, including will and desire, share the same biological substrate. As Levy points out:

[...] we can identify the self with a much broader network of states and mechanisms; if these states and mechanisms cause our actions, then *we* cause them²⁴.

In this sense, even consciousness, though determined by neural function, is an emergent property of that network. Thus, it cannot serve as a metaphysically independent instance capable of determining freedom, nor can it provide an exhaustive account of the self. Nevertheless, a sound account of consciousness can pave the way toward a possibility for free will. In this regard, a clarification is warranted: the discourse on consciousness is

²³ Similarly to the Cartesian’s theatre concept. Cfr. N. Levy, *Neuroethics: Challenges for the 21st century*, Cambridge University Press, Cambridge 2007, p. 25.

²⁴ *Ibid.*, p. 233.

exceedingly broad and intricate, meriting an independent, dedicated examination; thus, the forthcoming analysis will consider only the insights pertinent to the present discussion.

A compelling argument in this regard is presented by Michael F. Mascolo and Eeva Kallio: they propose an *embodied emergence*²⁵ model of consciousness, which clearly articulates all the observations made thus far. Consciousness, according to this account, results from a hierarchical integration of lower-order parts, leading to a higher-level organization. In the case of consciousness, biological subsystems, such as neural networks, integrate into a higher-order biological structure. This process involves at least two main characteristics: emergent systems exhibit new irreducible properties; these properties are not intended to be autonomous from the lower-level components. That is, the properties of emergent systems cannot acquire the ability to «override those of its parts»²⁶. If that were the case, Mascolo and Kallio suggest, the argument would revert back to a mind-body dualism. To further embodied emergence, they add that:

(2) A psychological process is any organismic process that mediates or is mediated by phenomenal experience and meaning, where meaning consists of the schematization of phenomenal experience. [...] we avoid characterizing psychological processes as "mental"—a term that can suggest something other than a biophysical process. What makes a biological process also a psychological process is its role in the constitution and mediation of experience and meaning. (3) Although phenomenal experience and meaning are emergent, the capacity for regulation or control, in itself, is not an emergent psychological property. Biological regulation is an already existing property of biological systems. (4) As emergent processes, psychological processes allow organisms to operate on the basis of the meaning and significance of their circumstances[...]. We suggest that consciousness, as intentional awareness, functions to integrate

²⁵ M.F. Mascolo, E. Kallio, *Beyond free will: The embodied emergence of conscious agency*, in «Philosophical Psychology», 32, 2019, p. 443.

²⁶ *Ibid.*, p. 444.

novel sources of information and plays an orienting rather than directing role in production of adaptive action. (5) Psychological processes, including consciousness, thus transform existing capacities for biological regulation[...] they are able to play a causal (but not autonomous) role in the production of action²⁷.

To enhance the present introduction regarding consciousness, it may be beneficial to incorporate a nuanced perspective on the conceptualization of conscious states. The debate surrounding this topic has yielded two opposing viewpoints: one perspective, namely psychophysicalism, is reductionist and draws upon neuroscientific data. The second viewpoint is that of functionalism, which is grounded in an antireductionist account of conscious events, referencing psychological inquiry²⁸. Jesse Prinz introduces the *neurofunctionalism*²⁹ approach, which is based on the constructive integration of both the aforementioned positions. According to Prinz, conscious states must be regarded as both functional and neural-based. This because—aligning with the embodied emergence previously assumed—on one hand, the level of neurons grounds the possibility for the psychological functions of consciousness; on the other hand, these psychological functions define the activity of those specific neurons, prefixed to carry out their tasks. These observations disclose what Prinz terms a two-level view of conscious states, uniting the level of neurons—the "realization" level—with the resulting psychological role, which is necessary in defining their functioning³⁰. Thus, consciousness cannot be properly *identified* with the self, but is nevertheless an important part of it, especially concerning free will.

²⁷ *Ibid.*, pp. 444-445.

²⁸ J. Prinz, *The conscious brain*, Oxford University Press, Oxford 2012, p. 286.

²⁹ *Ibidem*.

³⁰ Here Prinz excludes any form of teleology: neurons are not "programmed" to serve a particular final state, on the contrary this simply constitutes their natural way of functioning. Cfr. *ibid.*, pp. 291-293.

Furthermore, according to Prinz, consciousness arises when perceptual representations at the *intermediate level*³¹ are modulated by attention and made accessible to working memory. Prinz emphasizes that, although the brain processes numerous sensory signals, only those representations formed at the intermediate level and amplified by attention become part of conscious awareness. This occurs through synchronized neural activity, which integrates and stabilizes perceptual content into a unified experience. This conscious access enables the brain to evaluate options, predict outcomes, and make informed decisions. Thus, the interaction between attention, intermediate-level perception, and cognitive control is critical in decision-making processes.

To resume the initial question of the subsection: how, then, can the self be defined? It is the dynamic interplay of conscious and unconscious mechanisms, and it is not phenomenally detectable. As Prinz aptly writes:

There is no phenomenal I. If I wait for myself to appear in experience, I will never arrive³².

The self does not manifest as a distinct and consistent entity. Even if counterintuitive, the self is an adaptive illusion that supports coherence and interaction with the environment. As Corbellini and Sirgiovanni point out, this illusion is advantageous for survival, enabling stable perception and social functioning³³. Rather, what we perceive as the self is a byproduct of various perceptual and cognitive processes operating in concert. In this way, the self can be inferred indirectly by the fact that we experience the world through a specific point of view, beyond which we are unable to perceive: there are sensory boundaries³⁴.

³¹ This term refers to a stage of perceptual processing between low-level sensory detection (such as colors, shapes, or sounds) and high-level cognitive interpretation (such as identifying objects or understanding concepts). For further details cfr. *ibid.*, pp. 49-78.

³² *Ibid.*, p. 240.

³³ G. Corbellini, E. Sirgiovanni, *op. cit.*, p. 66.

³⁴ *Ibidem.*

An interesting study by Hanna and Antonio Damasio in this regard examines the role of homeostatic feelings³⁵. These feelings³⁶ link the body's internal milieu to the mind, establishing a foundation for conscious experience, personal perspective, and the regulation of life events. According to the neuroscientists, homeostatic feelings arise from the collaboration between the non-neural interior of the body and the interoceptive nervous system, which is significantly distinct from the exteroceptive or proprioceptive functions³⁷. Ultimately, this same set of processes serves as the filter through which our experience of the world is shaped, both at the sensory and cognitive levels:

So the senses do not simply pick up the world as it is; they impose order that would be invisible if we were different kind of beings. Our goals, interests and histories can contribute to this process of construction³⁸.

Having clarified the conception of self and consciousness, we can now explore their relevance to free will. A comprehensive understanding of the established processes driving decision-making may, perhaps, provide insight into the possibility of free will.

b. *The role of conscious deliberation*

Neil Levy offers an illuminating account of conscious deliberation that aligns with the reason-responsiveness model discussed earlier. Indeed, he acknowledges that humans are determined beings—our perceptions and actions emerge from sub-personal mechanisms that define how we process the world. Thus, we lack arbitrary control over decisions, a limitation he calls

³⁵ A. Damasio, H. Damasio, *Homeostatic feelings and the emergence of consciousness*, in «Journal of Cognitive Neuroscience», 36, 2024, pp. 1653–1659.

³⁶ Such as hunger and thirst, body temperature, or even nausea and pain. For further specifications, cfr. *ibid.*, p. 1654.

³⁷ The interoceptive neurons have specific characteristics that makes possible a more direct interaction with non-neural interior of the body, such as non-myelinated axons or the absence of blood-brain barrier (BBB). Cfr. *ibid.*, p. 1656.

³⁸ J. Prinz, *op. cit.*, p. 240.

the *decision constraint*³⁹, and he adds that possessing it would undermine, not enable, free will. Freedom instead arises in the passive process of *weighing* reasons instead of *weighting*⁴⁰. That is, evaluating, rather than arbitrarily altering, the significance they hold for the self is what truly matters. We are defined by all the conscious and unconscious mechanisms of the self; conscious deliberation, therefore, allows for the recognition of the reasons that matter most to us, rather than a detachment from causal determination.

For Levy, consciousness is essential because it integrates information into a *global workspace*⁴¹ that makes it accessible to multiple «parts of the brain simultaneously»⁴².

An analogous prerequisite regarding the importance of conscious psychological factors is proposed by Nahmias. He defends compatibilism⁴³ in contrast to the reductionism found in some accounts, which threaten the folk intuition of mind-body dualism in justifying free will⁴⁴. Indeed, Nahmias argues that psychological variables exert genuine causal influence⁴⁵. Alongside Oisín Deery, he adopts an interventionist perspective; that is, he considers the causal source of a specific event—, in this case, a decision or an action—, and inquires whether the same outcome could have manifested in the same way if those causes were manipulated⁴⁶. Given the plasticity of certain neural structures, Nahmias points out that a specific psychological function «could be realized by different neural variables»⁴⁷ maintaining the

³⁹ N. Levy, *op. cit.*, p. 233.

⁴⁰ *Ibid.*, p. 234. It may recall the difference between acting on own free will and freedom of the will talking about Frankfurt's theory.

⁴¹ *Ibid.*, p. 240.

⁴² This conception relies on the Ned Block's account of access-consciousness. Cfr. *ibid.*

⁴³ Through an *interventionist* account. It will be defined shortly.

⁴⁴ Nahmias precisely refers to Joshua Greene and Johnatan Cohen. Cfr. E. Nahmias, *Your brain as the source of free will worth wanting...*, cit., p. 3.

⁴⁵ As previously mentioned, this statement is made under the assumption of neuronaturalism; hence, it does not undermine the notion of the neural basis of mental functioning.

⁴⁶ For a more detailed explanation, cfr. O. Deery, E. Nahmias, *Defeating manipulation arguments: Interventionist causation and compatibilist sourcehood*, in «Philosophical Studies», 174, 2017, pp. 1255–1276.

⁴⁷ E. Nahmias, *Your brain as the source of free will worth wanting...*, cit., p. 10.

outcome's *invariance*⁴⁸. Conversely, altering psychological variables changes outcomes. Thus, psychological functions possess *causal invariance* and can be viewed as primary sources of agency. This implies that deliberative reasoning holds causal priority. Psychological processes, though realized in the brain, preserve explanatory autonomy, allowing to exploit functions based in our brains:

That deliberative event has a stronger causal invariance relation with my decision than any variables in my distant past⁴⁹.

Decisions made during reasoning in the deliberative process assume a fundamental causal status for free choice. When evaluating Nahmias' considerations in light of Prinz's neurofunctionalism, psychological constructs are functional properties realized in specific neural substrates; thus, neither epiphenomenal nor dualistic. Conscious reasoning, especially when it is self-reflective, allows the mind to integrate information and simulate alternative courses of action without invoking indeterminism.

Empirical research reinforces this view. Liad Mudrik and colleagues have demonstrated that manipulating contextual and cognitive factors can alter decision outcomes, indicating that many decisions are unconsciously biased⁵⁰. Nevertheless, awareness of such influences enables individuals to recognize and, to some extent, regulate them.

This study can be related to Sapolsky's discourse on deterministic factors, as these experiments replicate the conditions under which most individuals operate: our fragmented and complex nature is intrinsically influenced by numerous uncontrollable factors, even beyond the laboratory setting. This implies that often, even when engaging in conscious deliberation, decisions are influenced by several variables. A specific education, family,

⁴⁸ Namely the stability of the $X \rightarrow Y$ relation under a range of interventions on X and, more strongly, across relevant changes in background conditions C . This concept is explained in detail in O. Deery, E. Nahmias, *Defeating manipulation arguments...*, cit, pp. 1261-1263.

⁴⁹ E. Nahmias, *Your brain as the source of free will worth wanting...*, cit, p. 11.

⁵⁰ L. Mudrik, I.G. Arie, Y. Amir, Y. Shir, P. Hieronymi, U. Maoz, T. O'Connor, A. Schurger, M. Vargas, T. Vierkant, A. Roskies, *Free will without consciousness?*, in «Trends in Cognitive Sciences», 26(7), 2022, pp. 555-566.

genetic predispositions, or states of fatigue could unknowingly condition both the decisions we make and the manner in which we weighed reasons. For instance, if a person is very tired and feels the urge to be rude to her partner—due to a minor mistake—she may recognize valid reasons to succumb to her inclination to yell and rebuke. She feels angry about a particular cause—the partner’s mistake—and acting out of that anger could be perceived as useful in preventing similar mistakes in the future. In this very simple example, my aim is to highlight that a lack of awareness regarding the underlying causes of a certain decision-making process could impede free choice. Understanding that her feelings were rooted in ego depletion—due to a particularly tough day—or from having grown up in a quarrelsome family environment—which is linked to a predisposition to act impulsively—may have facilitated her deliberation and potentially averted that response. Thus, acting for reasons with which one identifies should also encompass the reasons behind one’s decision-making process. This is the sense in which consciousness could play a significant role in the eligibility of free will. As Mudrik affirms:

Being conscious of our motivations, reasons, decisions, and actions appears to endow us with a kind of executive control that we lack when these are not conscious⁵¹.

In this regard, the deeper our understanding of the factors that shape our actions, the greater our capacity to control them—and, consequently, the freer we become. PAP is integrated as a fundamental element of free will: the ability to re-evaluate reasons through an understanding of biases and conditioning can enable the possibility of acting otherwise.

[...] the more aware of the factors affecting our decisions, the more we can implement strategies to control them, if we like, and, thus, the freer we are. The more ignorant or deluded we are about the causes of our actions, the less control and freedom we have⁵².

⁵¹ *Ibid.*, pp. 561-562.

⁵² *Ibid.*, p. 563.

The awareness of causes enhances evaluative depth: one can recognize underlying causes yet still choose to act otherwise, as when one knowingly decides to follow an emotional impulse. The essential point is that deliberation informed by awareness expands the range of evaluative possibilities. In the previous example, the subject could have chosen to evaluate the emotional outburst positively, perhaps leading to a liberating cry and a justification of her behavior to her partner. Consciousness, therefore, provides the platform through which deterministically grounded agents attain self-reflective choice.

The constructive integration of emotional and cognitive states here is fundamental: they are both essential for sustaining a balanced and fruitful relationship with our environment and with others⁵³.

c. The neuroscientific contribution: the COINTOB model and "reasoning about reasons"

Finally, how can these considerations be translated into neuroscientific knowledge? Even though there is still a long way to go to deeply understand the neural mechanisms underlying decisions and consciousness, some results and insights from research could be helpful in supporting the proposed view. Libet's experiment showed that, as Corbellini and Sirgiovanni point out, his denial of free will aligns more with a discussion on the role of consciousness than with free will⁵⁴. To clarify, his experiments show that when making decisions because brain mechanisms are already activated before we can notice them⁵⁵. One evidently problematic concept is the conscious veto, which in Libet's words represents the only escape from a predetermined destiny. Nonetheless, if *positive* consciousness has no influence on the causality of a

⁵³ The interplay between emotional and cognitive areas of the brain is been demonstrated and discussed in a specific branch of studies regarding moral judgments. Since it is a broad debate, a rich and exhaustive overview can be found in G. Corbellini, E. Sirgiovanni, *op. cit.*; S. Songhorian, *Etica e scienze cognitive*, Carocci, Roma 2020.

⁵⁴ *Ibid.*, p. 131.

⁵⁵ After Libet, other scholars repeated the same experiment more accurately, since he has also been criticized for methodological issues.

specific course of action, why should the situation change with a *negative* one⁵⁶? The inhibition of the impulse that overcomes our consciousness could, in turn, be the result of brain activity. Other experiments have been conducted to evaluate this idea; in particular, Eamonn Walsh and colleagues detected, using EEG, cerebral activity during the inhibition of actions⁵⁷. This confirmed the hypothesis; it is indeed reasonable to think that the neurological system implements strategic mechanisms capable of reconsidering a certain action, so that if it turns out to be inconvenient, there can be a correction⁵⁸. Thus, the idea of a conscious causal power over actions appears to be inconsistent with neuroscientific studies.

How, ultimately, can decision-making be reconsidered? Roskies—referring to Michael Shadlen’s experiments on monkeys’ decision-making—illustrates a specific dynamic involved in transforming sensory information into a decision to act as «the skeleton of a decision-making system»⁵⁹. In the experiments considered, monkeys had to report a judgment by moving their eyes (i.e. a saccade) on one of two visual targets, based on the movement of dots. The cortical recordings of this process revealed that two areas of the visual cortex, namely MT and MST, were sensitive to the motion of the dots in a particular direction. In MT, the firing of cells more sensitive to rightward movement was stronger when the dots’ signal indicated rightward motion, and conversely. MST cells, much like MT, were also responsive to the transient perceptual signal of moving dots, albeit with larger receptive fields⁶⁰. The decisive element for this decision-making inquiry has been the one of the lateral inter-parietal area (LIP). Indeed, neurons in this area fire during the «planning and execution of an eye movement»⁶¹, receiving constant

⁵⁶ This is in part taken by the punctual observations De Caro does in M. De Caro, *Libero arbitrio e Neuroscienze*, cit., p. 80.

⁵⁷ P. Haggard, *Neuroethics of free will*, in J. Illes, B.J. Sahakian (eds.), *Oxford handbook of neuroethics*, Oxford University Press, Oxford 2011, p. 533.

⁵⁸ As Patrick Haggard suggests. Cfr. *ibidem*

⁵⁹ A. Roskies, *Freedom, Neural Mechanism, and Consciousness*, in R. Baumeister et al. (eds.), *Free will and consciousness: How might they work?*, Oxford University Press, Oxford 2010, p. 160.

⁶⁰ *Ibid.*, p. 159.

⁶¹ *Ibidem*.

information from different direction responses. Another interesting aspect of LIP neurons is that they fire also in the absence of a perceptual stimulus, unlike MT and MST. LIP neurons seemed to accumulate evidence and fire until the execution of the saccade:

These neurons look as if they are computing an estimate of total information for motion in that direction throughout the course of the trial⁶².

This progression of the firing rate of neural activity introduces an important element in the decision-making field: the existence of a threshold above which the decision is taken. In this sense, LIP fires when accumulating evidence, and when the evidence pertaining to a specific direction increases until the reaching of the threshold, the eye movement is executed. Roskies adds that similar experiments have been repeated with new features—for instance, the manipulation of the expected reward—resulting in altered LIP activity⁶³. This shows a surprising factor, that is LIP neurons are able to integrate different sources of information in the decision-making process, thanks to the interconnection with other brain regions.

Roskies posits that, even if these experiments pertain to perceptual decisions, the same pattern could also elucidate more complex decision-making processes, specifically those typically associated with free will. She asserts that the evidence accumulated by LIP neurons can be imagined as «propositions coding states of affairs»⁶⁴.

She continues in saying that it is possible to imagine analogous neural networks that—through the same process of accumulating evidence and firing until the threshold— underlie more complex and abstract decisions. The scheme of this decision-making process is one that encompasses reason-responsiveness, since it is grounded in the accumulation of evidence toward or against a certain decision. As also neuroscientist Kobe Dessenner asserts, it is now widely found that:

⁶² *Ibidem*.

⁶³ *Ibid.*, p. 161.

⁶⁴ *Ibid.*, p. 162.

Many kinds of decisions can be well explained in terms of the accumulation of evidence for different alternatives over time⁶⁵.

In this regard, Roskies affirms that the role of consciousness, or awareness, of reasons is fundamental because it provides for the propositions that could motivate a certain action. Her final proposal is that consciousness is decisive for deliberation that could be considered free, because the awareness of reasons can provide new information to be *weighed*⁶⁶ during the deliberative process⁶⁷.

This observation could be related to what was previously said about the importance of the awareness of causes behind certain reasons for action. Consciousness plays this fundamental role in providing even more information to the weighing process that occurs in the brain. In this regard, the concept of attention mentioned when assessing Prinz's account comes back useful, because it even more clarifies the fact that attended consciousness serves to make information available for flexible behavior. Thus, attention to this kind of reasons can implement new elements in the deliberation, making it more informed and consequently freer.

The last, more recent contribution I would like to integrate into the discussion is the one described by Marcel Brass and colleagues. It could give a good account of the study analyzed by Roskies, which included the firing activity of neurons in relation to a threshold. In their article, they review the body of research that followed the renowned Libet experiment, tracing the various positions and criticisms that arose from its results⁶⁸.

They illustrate that new findings suggest the Libet experiment has not been correctly interpreted; one of the primary criticisms pertains to the

⁶⁵ K. Desender, K.R. Ridderinkhof, P.R. Murphy, *Understanding neural signals of post-decisional performance monitoring: An integrative review*, in «eLife», 10, 2021, p. 2.

⁶⁶ The actual term Roskies utilizes is "weighting", but here I would like to remark the specific semantic meaning Levy assigned to the word.

⁶⁷ A. Roskies, *Freedom, Neural Mechanism, and Consciousness*, in *Free will and consciousness*, cit., p. 167.

⁶⁸ There is a lot of interesting information regarding single experiments that could not be assessed here. For further details, cfr. M. Brass, A. Furstenberg, A. R. Mele, *Why neuroscience does not disprove free will*, in «Neuroscience & Biobehavioral Reviews», 102, 2019, pp. 251–263.

explanation of the RP. Indeed, Libet asserted that it coincides with the motor preparation process of an action already undertaken. Conversely, recent studies indicate that the RP is part of the decision-making process itself, which culminates in action when a threshold is surpassed. There are several factors that can influence this accumulation of neural activity, depending on the task. In situations similar to that of the Libet experiment, where the subject must choose when to do something, participants form conditional intentions relative to the task instruction. For example, choosing between options within a designated time frame. These conditions will influence the decision-making process, setting what type of information is accumulated and the threshold setting. Moreover, it has also been found that even after surpassing the threshold, it remains possible to veto the action for a short time window for the integration of post-decisional evidence accumulation.

They assert that—recalling Roskies’ considerations on Shadlen’s experiments— these types of decision processes occur in a manner akin to perceptual decision processes, whereby the RP indicates the accumulation. The sole distinction resides in the nature of the information that is accumulated to bound, that is the threshold. They propose a model that interprets the RP in this manner:

RP is a neural index of the continuous integration of information and stochastic neural activity⁶⁹.

They call this specific model for decision-making the conditional intention and integration to bound (COINTOB) model⁷⁰.

⁶⁹ *Ibid.*, p. 259.

⁷⁰ *Ibidem.*

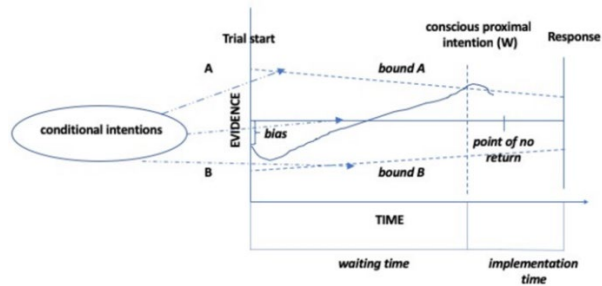


Figure 1. The COINTOB model.

Figure 1 provides a graphical illustration of a Libet-style experiment: the diagram depicts the process of information integration, beginning with a set of conditional intentions that shape the process. Furthermore, it can be observed that the irregular curve in the figure initially trends toward B; however, the accumulation of evidence subsequently surpasses the bound to A. W represents the moment in which participants become conscious of their choice, which may be influenced by new evidence prior to reaching the point of no return. The process culminates in the execution of the action. Brass and colleagues state that the only difference between this type of experiments and deliberation, as previously mentioned, is the kind of evidence integrated in the process. These kinds of decisions involve a process of «arbitrary picking decision»⁷¹; meaningful deliberation differs mainly in that the integration to bound is informed by internal evidence rather than stochastic noise⁷², which

⁷¹ *Ibidem*.

⁷² This particular phenomenon is still being studied, but Shadlen and Roskies declared that «As far as we understand, the existence of noise does not confer any special properties, like freedom, will, consciousness, etc.». Cfr. M. N. Shadlen, A. L. Roskies, *The neurobiology of decision-making and responsibility: reconciling mechanism and mindedness*, in «Frontiers in neuroscience», VI, 56, 2012, p. 6. Another interesting fact about this "background" noise, can be found in V. Axelrod, G. Rees, M. Lavidor, M. Bar, *Increasing propensity to mind-wander with transcranial direct current stimulation*, in «Proceedings of the National Academy of Sciences», 112, 2015, pp. 3314–3319. According to this study, it turned out to be useful for one particular purpose among the others: it is powered by the same dorso-lateral prefrontal cortex (dlPFC), for this state is linked to mind wandering and thus to creative mode and future planning, both advantageous functions.

is typical of decisions that lack significant evidence. Hence, the relevant point for the discussion on free will, according to Brass and colleagues, is this:

According to our ITB model of intentional choice, the decision-making process usually involves substantive information, and it relies on noise only when substantive information is not available⁷³.

This model further elucidates the possible dynamics underlying the deliberative process, suggesting that—since internal evidence contributes to reaching the bound—conscious awareness of causes underlying reasons to act could constitute a fundamental element for free will guided decisions. Moreover, the fact that veto is still contemplated in the process could imply that conscious attention can integrate new evidence even when the decision is taken, although timing within deliberative processes could sensibly differ compared with picking situations. Desender and colleagues considered in detail the role of post-decisional evidence in relation to the detection of an error and confidence by experimental subjects. As new evidence is accumulated, a second boundary is crossed in which a subject can signal an error in their evaluation⁷⁴. Considering conscious deliberation, the memory of the importance of causes behind reason for acting can constitute an important candidate for the decision⁷⁵.

4. Conclusion

The final paragraph aimed to investigate the scientific and philosophical debate in search of a possible solution for admitting free will in such a determined world. The solution that, arguably, seems to work is that of a compatibilist account of free will which relies on reason-responsiveness and so on possible alternative actions. The reason-responsiveness discussed herein refers to an awareness of a deeper layer of reasons, particularly the

⁷³ M. Brass, A. Furstenberg, A.R. Mele, *Why neuroscience does not disprove free will*, cit., p. 261.

⁷⁴ K. Desender, K. R. Ridderinkhof, and P. R. Murphy, *op. cit.*, p. 6.

⁷⁵ Desender points out that, among other factors, «new information that comes to mind from memory» is one of the qualitative sources of information that are available during veto, as intended throughout this paragraph. Cfr. *ibid.*, p. 6.

possible causes underlying the selection of a specific motivation. This awareness could facilitate a more informed evaluation, resulting in a freer choice of a particular course of action. Consequently, this could also enhance an individual's capacity for a more reflective and effective approach to problem-solving. Neuroscientific work on decision-making is a promising and growing body of knowledge for what concerns the free will debate. New research avenues and discoveries will continue to nourish the theoretical discussion, and in this context, my proposal could be enriched or disconfirmed.

The limitation of this perspective is that the ability to "reason about reasons" is not always attainable. Since unquantifiable factors shape our development, one can find himself immersed in an environment that hinders conscious awareness, and consequently the capacity to react to reasons during the decision-making process. This point may be well developed through a focus on the possibility of cognitive and moral enhancement. Research may open not just to direct means of enhancement—such as pharmacological and neuro-engineering ones—but also, and foremost, to environmental and educational stimulus which could train and improve capacities fundamental for recognizing the aforementioned way of deciding.