An Introduction to the Hard Problem of Consciousness

Part 1: From Descartes' Cogito to Chalmers' Zombies: A Historical Introduction to the Hard Problem of Consciousness

---

This two-part series is intended as an accessible introduction to one of the central problems in contemporary philosophy of mind: what David Chalmers called the "hard problem of consciousness." Rather than immediately offering an abstract definition, I have chosen to retrace its genesis. The hard problem belongs to a long history of attempts to understand the place of mind and consciousness in the physical world, and in particular the relation between mind and brain.

I will therefore begin with Cartesian dualism and the problem of interaction between mind and body, before following the successive transformations of this question: the rise of materialism and then physicalism, the development of physiology and neuroscience, the principle of causal closure, identity theory, and functionalism. This trajectory will help us understand why, even after the physical and functional explanation of behavior has made considerable progress, one difficulty seems to remain: why are these processes accompanied by subjective experience?

The aim of this first article is therefore both historical and conceptual: to show how the hard problem emerged and to arrive gradually at its contemporary formulation. A second article will be devoted to the main responses that have been proposed and to their respective costs. I will not pretend to adopt an artificially neutral stance there: some solutions seem more convincing to me than others, but their comparison will come only after the problem itself has been clearly stated.

I. A Convenient Point of Entry: Descartes

In his Meditations on First Philosophy, Descartes develops, in several stages, an argument intended to establish that the mind is a substance distinct from matter. [Descartes (1641)]

He begins, in Meditation II, with methodological doubt: I can doubt everything, including the existence of my body, but I cannot doubt that I am thinking while I doubt. This certainty is not the product of deductive reasoning—Descartes insists on its immediate character: "I think, therefore I am" presents itself to the mind through direct intuition, without relying on a prior general premise. This is the cogito. Descartes concludes that he is first and foremost a thinking thing—a mind, an intellect—while at this stage avoiding the word "soul," too closely associated with the scholastic vocabulary from which he is trying to distance himself.

This first conclusion is not yet enough to establish a real distinction between thought and matter. Being able to conceive that my body does not exist does not guarantee that it is genuinely possible for my thought to exist separately from it: there is a gap to bridge between conceivability and real possibility. This is the task of Meditations III to V, in which Descartes establishes the existence of a non-deceiving God: this guarantee is supposed to ensure that whatever is conceived clearly and distinctly corresponds to a genuine possibility, since God can create anything that is conceived in this way.

This divine guarantee (quite apart from the fact that it rests on the highly controversial ontological proof of God's existence) nevertheless raises a famous difficulty, formulated in particular by Antoine Arnauld in the Fourth Objections: does Descartes not reason in a circle? To prove the existence of a non-deceiving God, he relies on arguments that he takes to be clear and distinct; yet it is then the existence of this non-deceiving God that is supposed to guarantee that whatever he clearly and distinctly perceives is true. The rule of clear and distinct ideas therefore seems to be used before its reliability has been established. Descartes replies that a clear and distinct perception carries certainty as long as it is actually present to the mind, and that the divine guarantee is needed only to secure the lasting reliability of memory. This reply avoids the crudest form of the circle, but it remains disputed: the argument for God itself seems to require a chain of reasoning whose reliability cannot be continuously sustained by a presently occurring intuition. [Arnauld (1641); Descartes (1641)]

Only then, in Meditation VI, can Descartes conclude: since he clearly and distinctly conceives his mind as wholly separable from his body, and since such clear and distinct conception guarantees a real possibility, thought and matter are genuinely distinct substances. This is Cartesian dualism.

This dualism was not, however, an obvious starting point. Thomas Hobbes, a contemporary of Descartes and the author of the Third Objections to the Meditations in 1641, was already defending the opposite materialist position: for him, everything that exists, including what we call thought or will, is nothing but body in motion, with no need to posit a second substance distinct from matter. [Hobbes (1641)]

This dualism nevertheless creates a problem, identified by Princess Elisabeth of Bohemia in her correspondence with Descartes: how can two substances of different natures interact causally, so that our thought is affected by what happens to our body and our body is moved by our thought? Descartes first proposes a physiological site for this interaction, the pineal gland, but this answer does not touch the heart of the conceptual problem raised by Elisabeth. Vaguely identifying a gland tells us nothing about the mechanism by which an unextended substance could move an extended one, and Descartes eventually changes terrain rather than answering Elisabeth directly. He concedes that the union of soul and body cannot be conceived with the same tools by which they are distinguished: the distinction belongs to pure understanding, whereas the union is grasped through ordinary life and the senses—hunger, pain, and desire provide a direct experience of it, not a demonstration. He even suggests that excessive metaphysical reflection on this union can obscure what we already know confusedly through the simple fact of embodied life. This shift is not a straightforward admission of failure—Descartes maintains that his doctrine needs no correction, and that it is the demand for rational clarity that is misplaced when applied to this object. But he never provides a causal mechanism, and Elisabeth never declares herself fully convinced: their exchange continues through their correspondence and later feeds into The Passions of the Soul. The difficulty she brought to light has remained, in the eyes of most commentators, one of the points at which Cartesian interactionist dualism displays its clearest fragility. [Elisabeth of Bohemia & Descartes (1643)]

II. Three Responses to the Challenge of Interactionism

This problem of interaction, left open by Descartes, would structure a large part of seventeenth-century metaphysics: each of the major successors to Cartesianism would try to remove it by rethinking, in his own way, the relation between thought and matter.

Spinoza takes the most radical route: he eliminates the problem by eliminating its condition of possibility, namely the plurality of substances. In the Ethics, there is only one infinite substance, which Spinoza calls indifferently God or Nature. This single substance has infinitely many attributes[1], of which only two are known to us: thought and extension[2]. In other words, what we call "my body" and "my mind" do not correspond, for Spinoza, to two different things that must somehow communicate. They are one and the same reality considered in two different ways: as extended reality when described under the attribute of extension, and as mental reality when described under the attribute of thought. Hence the famous parallelism: the order and connection of ideas is the same as the order and connection of things. There is never causation from mind to body or from body to mind, because there are never two separate substances that need to communicate—only one reality viewed from two angles. [Spinoza (1677)]

This position anticipates both certain forms of neutral monism and what would later be called "dual-aspect" theories, two families whose boundary is not always clear. Monism here means that, at the fundamental level, there is only one reality or substance. In neutral monism, this fundamental reality is in itself neither specifically material nor specifically mental: the mental and the physical are then two ways in which that reality can manifest itself, be organized, or be described.

Malebranche, while remaining within the Cartesian framework of two distinct substances, solves the problem by relocating it to God. For him, no creature—neither body nor mind—is genuinely the cause of anything: only God possesses true causal power. When my will seems to move my arm, or when a burn seems to produce pain, these are in fact "occasions": my mental or bodily state is merely the occasion on which God, according to laws he himself has established, produces the corresponding effect in the other substance. This is the

[1] In Spinoza, an attribute is a fundamental way in which the intellect apprehends the essence of substance. Thought and extension are therefore not two parts of God or Nature, but two irreducible ways in which the same reality can be understood.

[2] In this context, extension refers to the bodily or physical domain: that which occupies space.

doctrine of occasionalism. It avoids any direct interaction between the two substances, at the cost of making God the permanent and unique causal agent of everything that happens in the world. [Malebranche (1674–1675)]

Leibniz, finally, rejects both Cartesian interactionism and Malebranche's occasionalism, which he regards as unsatisfactory because it requires continual divine intervention—what he calls a "perpetual miracle." His metaphysics is based on monads, simple substances without parts and, above all, "without windows": nothing enters or leaves a monad, and each unfolds its internal states according to its own internal principle, independently of all the others. If mind and body nevertheless appear coordinated—if I want to raise my arm and my arm rises—it is not because one acts on the other, but because God arranged, once and for all at creation, the unfolding of every monad so that they perfectly agree with one another, like two clocks synchronized from the beginning and never needing to be adjusted again. This is pre-established harmony: a single initial miracle replaces the continual interventions required by occasionalism. [Leibniz (1695, 1714)]

Leibniz also formulates an objection to what we might call a mechanistic explanation of mind: the idea that thought and perception could be fully explained by the organization, movements, and interactions of the different parts of a material system. This objection strikingly anticipates some contemporary debates about consciousness. In section 17 of the Monadology, he imagines a machine whose organization is capable of producing thought, sensation, and perception. Suppose this machine were enlarged while preserving its proportions until one could enter it as one enters a mill. Walking through its interior, one would find only parts acting on one another, without discovering anything that could explain perception itself. [Leibniz (1714)]

This mechanistic conception also corresponds to an important form of classical materialism: it is already found in Hobbes and later, in different forms, in La Mettrie and d'Holbach—although Diderot moves away from it by defending a more dynamic conception of matter. We will return to these authors in the next section.

Leibniz does not merely conclude that our mechanical explanation of thought and perception is provisionally incomplete. He argues that perception is, in principle, inexplicable through figures and motions alone, and must therefore be sought in a simple substance rather than in a composite or a machine. The mill argument is thus directly embedded in his theory of monads. Nevertheless, it anticipates an intuition that will reappear in a different form in the hard problem of consciousness: even an exhaustive description of the mechanisms and interactions among the parts seems to leave open the question of why, or even how, lived experience results from them.

III. The Materialist Response to the Cartesian Problem

The sequence that follows should not be read as a direct debate with Leibniz. There is no indication that La Mettrie or d'Holbach constructed their materialism as an explicit reply to the mill argument. Their main opponent remained Cartesian dualism. They do reject Leibniz's conclusion by maintaining that thought and sensation can belong to an entirely material organism, but they do not seem to confront directly the specific difficulty he raises: how can the organization and interactions of the parts of a machine account for the very existence of lived perception? On this point, the problem largely remains open.

La Mettrie leads the way with L'Homme Machine (1747), taking to its logical conclusion the animal-machine that Descartes had reserved for non-human animals: the human being is itself nothing but an organized machine, and thought is nothing more than a function of this material organization, with no need to posit a separate soul or a site of interaction. [La Mettrie (1747)]

Diderot begins from a more cautious position—close, in his early writings, to classical deism—before moving toward a fully developed materialism, notably in Le Rêve de d'Alembert. There he defends the idea that sensitivity is not the privilege of an immaterial soul but a general property of matter itself, sometimes inert and sometimes active depending on its organization—a less mechanistic solution than La Mettrie's, closer to a kind of materialist vitalism. [Diderot (1769)]

D'Holbach, finally, gives this movement its most systematic formulation in Système de la nature (1770): nature is described as a material whole governed solely by its internal laws, with no need for any spiritual or divine cause to account for it. The mind-body interaction problem that preoccupied Descartes' immediate heirs is therefore not solved but dissolved: for these authors, there is now only one substance to explain. [d'Holbach (1770)]

IV. The Nineteenth Century: Three Empirical Constraints

The nineteenth century does not directly settle the debate between dualists and materialists, but it accumulates empirical constraints that make the dualist position increasingly difficult to maintain and prepare the ground on which, in the following century, the idea of the causal closure of the physical will be formulated. We will return to that idea later.

The first constraint comes from Darwinism. If the human mind belongs to a gradual continuum of mental complexity, from the simplest organism to human beings, with no sharp break in the evolutionary lineage, it becomes difficult to justify the precise moment at which an immaterial soul would have been added to matter. This is not a causal argument like the next two: Darwinism challenges less the mechanism of mind-body interaction than the metaphysical rationale for the emergence, in the human species alone, of a substance of a radically different kind. [Darwin (1871)]

The second constraint comes from the development of classical mechanics and the laws of conservation, formulated and systematized during the nineteenth century notably by Mayer, Joule, and Helmholtz. Conservation of energy requires that, in an isolated physical system, the total amount of energy remain constant. At first sight, the intervention of an immaterial substance in the operation of the body therefore seems problematic: if the mind produced movement by adding a cause external to physical dynamics, one might expect an unexplained change in the energy balance.

Conservation of momentum adds a further difficulty. Supposing that the mind creates no energy but merely redistributes it or redirects motion does not necessarily save interactionism: such a deviation could itself alter the total momentum of the system. To these two constraints is added the deterministic ideal of classical mechanics, famously expressed by Laplace: if the complete physical state of the world at one moment, together with the laws governing it, entirely determines its future evolution, there seems to be no physical event whose outcome an immaterial will could modify without introducing a break in that evolution.

It is important, however, not to confuse these different constraints with the stronger principle of the causal closure of the physical, according to which every physical effect has a sufficient physical cause. The conservation laws and classical determinism were certainly conceived as very general principles of the physical domain, but their content alone does not state that every cause of a physical event must itself be physical. An interactionist dualist can precisely deny that brain processes form a causally complete system governed by these laws alone when a mental cause intervenes. The dualist may maintain that, at certain moments in the interaction between mind and brain, physical evolution is not entirely determined by the prior physical state, or even that the conservation relations usually observed do not hold without exception.

Classical physics therefore did not logically refute interactionist dualism. It did, however, substantially increase its explanatory burden. The more precisely physical and physiological processes are described without any point being observed at which their evolution requires the intervention of a non-physical cause, the more the dualist must specify where that intervention occurs, how it alters brain dynamics, and why its effects have not yet been detected. The logical possibility of an exception remains, but it gradually becomes an empirical hypothesis that is especially difficult to establish rather than an idea that flows naturally from our best scientific explanations of how the world works.

The principle of causal closure generalizes precisely this development: it states that there is no such point of openness—in other words, that physical causes are sufficient to complete the explanation of every physical event. This is why adopting the principle places a much more direct constraint on interactionist dualism than either the conservation laws or classical determinism considered in isolation.

This latter constraint would, however, lose some of its force with the advent of quantum mechanics, some interpretations of which introduce outcomes that are only probabilistically determined. This opening is not enough to solve the interaction problem: replacing a determined evolution with several possible outcomes still does not explain how a non-physical will could select one of them. We will return to this point.

The third constraint comes from experimental physiology and then from the emerging field of neurology. Around 1850, Hermann von Helmholtz succeeded in measuring the propagation speed of the nerve impulse. The result does not directly refute dualism: nothing yet prevents its defender from supposing that an interaction with mind occurs at some particular point in the neural chain. But it contributes to a decisive shift. Phenomena once associated with the immediate action of a vital or mental principle—the transmission of sensation, the preparation of movement, the delay between stimulation and response—become quantifiable physical processes embedded in a measurable physiological timescale.

The work of Paul Broca, beginning in 1861, and then of Carl Wernicke in the 1870s strengthens this constraint by establishing systematic relations between certain brain lesions and specific language deficits. Damage to frontal regions of the left hemisphere can, for example, severely impair articulated speech production, whereas other lesions affect comprehension more strongly. These results show neither that every mental faculty has a single, sharply delimited location nor that the mind reduces to the brain. They do show, however, that specific mental capacities depend on the integrity and organization of particular brain structures. The dualist can therefore no longer be satisfied with asserting a general dependence of mind on body: they must explain why local alteration of physical tissue produces relatively specific mental deficits.

It was the physiologist and philosopher Thomas Huxley who, in 1874, drew one of the most radical consequences from this progressive naturalization of the nervous system. In "On the Hypothesis that Animals Are Automata, and Its History," he defends the hypothesis that animals, human beings included, are conscious automata: their movements and behavior can be fully produced by nervous mechanisms, while their conscious experiences accompany those processes without exerting any causal influence of their own. Huxley thus compares consciousness to the steam whistle of a locomotive, which signals the operation of the machine without contributing to its motion. [Huxley (1874)]

This position would later be called epiphenomenalism: brain events cause conscious experiences, but those experiences produce no physical effects in return. We should not, however, conclude that Huxley necessarily reintroduces a soul or a separate mental substance. His claim is primarily about the causal inefficacy of consciousness, without offering a fully developed theory of its ontological status[3]. It nevertheless occupies a pivotal position: Huxley accepts that consciousness genuinely exists and is linked to brain function, while abandoning the idea that it plays a role in producing behavior.

One final milestone deserves mention because it directly touches what would later become the hard problem of consciousness. In 1872, the physiologist Emil du Bois-Reymond delivered a lecture entitled "On the Limits of Our Knowledge of Nature." He argued that even if physiology were one day able to describe the physical processes of the brain in full, it still could not thereby explain how conscious experience could arise from motions of matter. The difficulty, in his view, is not merely the temporary insufficiency of our knowledge: no arrangement of particles and no mechanical description can make intelligible the transition from neural activity to lived sensation. [du Bois-Reymond (1872)]

Du Bois-Reymond summarizes this limit with the phrase that became famous: Ignorabimus—"we shall not know." His position is nevertheless more cautious than straightforward dualism. A mechanistic physiologist and opponent of vitalism, he holds that

[3] Ontology concerns what exists, what is real. Here, Huxley is not discussing what consciousness really is.

mental phenomena have material conditions and must be studied scientifically. But he refuses to conclude that their nature is wholly reducible to matter, whose ultimate essence he also regards as unknown. He therefore defends less a precise ontology than a form of metaphysical skepticism: the natural sciences may discover the correlations and physical mechanisms associated with consciousness without necessarily making the very existence of subjective experience intelligible. A formulation very close to the later explanatory gap thus appears as early as 1872, more than a century before Chalmers gave the problem its current name.

This attitude anticipates what would much later be called mysterianism: the idea that certain questions, especially the relation between brain and conscious experience, might in principle lie beyond the cognitive capacities of human beings. Contemporary defenders of this position do not necessarily claim that the problem is insoluble in itself, but rather that our minds may lack the conceptual resources needed to solve it. [McGinn (1989)]

This difficulty also nourished an intermediate route between traditional dualism and strictly reductive materialism: British emergentism. As early as 1843, John Stuart Mill distinguished merely compounded effects, which can be understood by adding together the effects of their causes taken separately, from effects he called “heteropathic,” in which genuinely new properties or regularities appear. The properties of water, for example, do not resemble a simple addition of the properties of the hydrogen and oxygen that compose it: their combination produces a system with new characteristics. George Henry Lewes later introduced the term emergence for this kind of novelty, before Samuel Alexander, C. Lloyd Morgan, and C. D. Broad systematized the position in the early twentieth century. [Mill (1843); Lewes (1875); Alexander (1920); Morgan (1923); Broad (1925)]

Applied to mind, emergentism holds that consciousness is not a substance added to matter from the outside: it appears when physical organization reaches a certain degree of complexity. Conscious properties would nevertheless not be deducible from the laws governing elementary physical constituents alone. This strategy preserves the mind’s dependence on the brain without accepting a complete reduction of the former to the latter. It thus anticipates a family of views that will repeatedly reappear in contemporary debates: consciousness would be wholly natural and produced by the physical, while still constituting a level of reality with properties or laws of its own.4

V. From Vitalism to Modern Neuroscience: The Naturalization of Life and Mind

This section deliberately follows a thematic rather than a strictly chronological thread. From the turn of the twentieth century into its second half, several experimental programs converge in making living systems, the nervous system, and a growing range of mental functions accessible to physicochemical and physiological explanation—not without one

[4] See also my article [Vila Valls (2026)], “Émergence, réduction et real patterns,” The Naturalistic Outlook, https://thenaturalisticoutlook.com/2026/07/19/emergence-reduction-et-real-patterns/

final resurgence of vitalism[5] along the way. This progression should not, however, be confused with a general conversion of researchers to materialism: several of them would help extend the domain of naturalistic explanation while retaining dualist or non-reductive metaphysical positions.

The first advance concerns the very structure of the nervous system. In 1873, Camillo Golgi developed a staining technique that made it possible to visualize individual nerve cells, but he interpreted his own results in favor of a reticular theory according to which the nervous system formed a continuous network. Santiago Ramón y Cajal, using the same technique, instead established the neuron doctrine: each nerve cell is a distinct anatomical unit, separate from the others and communicating with them without direct physical continuity. Golgi and Cajal nevertheless received the Nobel Prize jointly in 1906, while each continued to defend his own interpretation.

This discovery paved the way for the work of Charles Sherrington, who introduced the term “synapse” at the end of the nineteenth century to designate the functional zone of communication between two neurons. In The Integrative Action of the Nervous System (1906), he showed how elementary reflexes could be coordinated and integrated by the organization of the nervous system: the physiological explanation of this coordination could be sought in the properties of neural circuits and synaptic interactions themselves. [Sherrington (1906)] We should not, however, infer from this that Sherrington was a materialist. In his later reflections on the relation between mind and brain, notably in Man on His Nature (1940), he remained reluctant simply to identify the mental with the neural and retained a dualist—or at the very least strongly non-reductive—conception. [Sherrington (1940)] His case therefore illustrates a distinction that will be essential for what follows: a scientific program can naturalize a function without its author concluding that all mental reality reduces to the physical.

The movement toward naturalization was not linear. At the same time, the embryologist Hans Driesch revived vitalism on serious scientific grounds. His experiments on sea urchins, beginning in 1891, showed that when the first two cells of an embryo were separated, each could develop into a complete and normal organism—a result that the mechanistic models then available found difficult to explain. In his Gifford Lectures of 1907–1908, Driesch concluded that there must be an “entelechy”: a non-physical organizing principle specific to living things, supposedly guiding the development of the organism without being reducible to the physicochemical mechanisms composing it. [Driesch (1908)]

It was in chemistry and experimental biology that this revived vitalism encountered its most direct opponents. As early as 1897, Eduard Buchner showed that alcoholic fermentation could occur in a simple cell-free yeast extract, without any intact living cell being required. Jacques Loeb extended this anti-vitalist program with his experiments on artificial parthenogenesis in sea urchins from 1899 onward: through physicochemical manipulations,

[5] Vitalism is the idea that the phenomena characteristic of living things cannot be fully explained by physicochemical laws alone and require a principle specific to life, sometimes conceived as a force or organizing principle. In various forms, this way of thinking was widespread before physiology, biochemistry, and biology progressively succeeded in explaining a growing number of vital phenomena without appealing to such a principle.

he triggered the development of unfertilized eggs into normal larvae. In The Mechanistic Conception of Life (1912), he explicitly defended the goal of a physicochemical explanation of vital phenomena. [Buchner (1897); Loeb (1899, 1912)] His approach is also associated with what the historian Philip Pauly called an “engineering standpoint”: to understand living systems is also to acquire a degree of experimental control over their mechanisms comparable to that sought by engineers in artificial systems. [Pauly (1987)]

This program took on a new dimension with the rise of molecular biology. In 1944, Oswald Avery, Colin MacLeod, and Maclyn McCarty showed that the “transforming principle” capable of transmitting certain hereditary traits in pneumococci consisted of DNA. Heredity thus began to be linked to a specific chemical substance rather than to an irreducible biological principle. [Avery, MacLeod & McCarty (1944)]

In 1953, James Watson and Francis Crick proposed the double-helical structure of DNA, drawing in particular on X-ray diffraction data obtained at King’s College by Rosalind Franklin and Raymond Gosling, as well as by Maurice Wilkins and his collaborators. The complementarity of the bases immediately suggested the outline of a copying mechanism: if the two strands separate, each can serve as a template for reconstructing its complement. A central phenomenon of life—the transmission of hereditary information—thus became intelligible through the structure and interactions of a molecule. [Watson & Crick (1953); Franklin & Gosling (1953); Wilkins, Stokes & Wilson (1953)]

The continuity with Loeb’s program is primarily intellectual rather than genealogical: molecular biology spectacularly realizes the ambition of making vital phenomena manipulable and explicable through their mechanisms. Jacques Monod particularly well embodies this turn. With François Jacob, he helped elucidate molecular mechanisms regulating gene expression in the early 1960s. [Jacob & Monod (1961)] Then, in Chance and Necessity (1970), he gave this biology an explicitly naturalistic philosophical interpretation: the apparent purposiveness of organisms—their “teleonomy”—must be understood through molecular mechanisms, heredity, and natural selection, not as the manifestation of an external plan or a vital principle. [Monod (1970)]

A parallel transformation took place in the sciences of the brain. Edgar Adrian’s work, made possible by improvements in electrical amplification techniques, allowed researchers in the 1920s to follow the activity of individual nerve fibers. It showed in particular that a neuron produces elementary electrical impulses of roughly similar amplitude: a stronger stimulus therefore does not simply produce a much stronger impulse. Instead, it can be represented by a greater number of impulses emitted per second, as well as by the recruitment of a larger number of nerve fibers. Very roughly, one can compare this with Morse code: information is not carried by the intensity of a single “beep,” but by the way signals are distributed over time. In the nervous system, information therefore depends not simply on the strength of each impulse, but notably on their frequency and on the number of fibers participating in the response. The nervous system thus begins to be described not merely as an anatomical network, but as a device capable of physically transmitting and encoding information. [Adrian (1932)]

In 1952, Alan Hodgkin and Andrew Huxley took a further step by explaining the physical mechanism of these nerve impulses, known as action potentials. The membrane surrounding a neuron separates environments containing different concentrations of ions—especially sodium and potassium, that is, electrically charged particles. When a neuron is activated, certain passages in its membrane open and close in succession, allowing these ions to flow. These movements of electrical charge produce the rapid change in voltage that constitutes the action potential and allow it to propagate along the neuron. The impulse whose speed Helmholtz had measured a century earlier thus received a precise electrochemical description. [Hodgkin & Huxley (1952)]

With David Hubel and Torsten Wiesel, physiological explanation then reached the processing of perceptual information itself. Their work on the visual cortex showed that certain neurons respond selectively to specific features of stimuli, notably the orientation and position of edges, and that the organization of cortical circuits contributes to the progressive analysis of visual information. [Hubel & Wiesel (1962)] The task was no longer merely to explain how a nerve signal propagates, but how the activity of populations of neurons can perform specific perceptual operations.

Roger Sperry's research on so-called split-brain patients further strengthened this functional dependence of the mental on brain organization. In some patients with severe epilepsy, surgeons had severed the corpus callosum—the large bundle of nerve fibers connecting the two cerebral hemispheres—sometimes together with other interhemispheric connections. The two halves of the brain could therefore exchange certain kinds of information much less easily. Experiments then showed that information received primarily by one hemisphere could become inaccessible to the other, and that cognitive capacities previously functioning as a unified whole could dissociate in systematic ways. [Sperry (1968)] A material alteration of brain connections therefore does not merely produce a global deficit: it transforms, in a structured way, the organization of the mental functions available to the subject.

By the middle of the twentieth century, two initially distinct programs thus converged. Molecular biology progressively made heredity, regulation, and a growing share of cellular functioning intelligible through physicochemical mechanisms; neuroscience, in parallel, decomposed the transmission, encoding, and processing of neural information into electrical, chemical, and anatomical mechanisms. None of this by itself demonstrates that consciousness is reducible to the physical: Sherrington, and later Eccles, show precisely that one can accept much of this scientific naturalization while remaining a dualist. But the room left for an independent mental intervention narrows as functions once attributed globally to the "mind" become accessible to detailed physiological explanation.

This continuity of program would appear especially strikingly in Francis Crick: one of the architects of the molecular revolution in DNA would devote The Astonishing Hypothesis (1994), forty years later, to the idea that perception, thought, and the sense of personal identity should likewise be approached as products of brain activity. [Crick (1994)] The success of this extension is precisely what contemporary debates about consciousness will have to determine.

VI. The Quantum Interlude: A Door Ajar, Then Closed Again

The previous section followed the development of biology and neuroscience into the second half of the twentieth century. We must now go back a few decades, to the 1920s, because a revolution in physics would briefly seem to reopen a door that this progressive naturalization appeared to be closing. The advent of quantum mechanics in the 1920s challenged the deterministic framework of classical physics, at least in the interpretations that quickly became dominant. The Born rule, which allows experimental predictions to be extracted from the theory's mathematical formalism, assigns only probabilities to the different possible outcomes and, in the orthodox theory of quantum measurement, no element of the quantum state determines which one will occur in a particular observation. Even complete knowledge of the physical state would therefore no longer make it possible to predict every individual event with certainty: Laplacian determinism appears to have been abandoned. [Born (1926)]

This conclusion nevertheless depends on the interpretation adopted. The Schrödinger equation, which is the central equation describing the evolution of the quantum state, is itself continuous and deterministic. In the orthodox formulation, however, a measurement is assumed to introduce a break in that evolution: among the different possible outcomes, only one is obtained, with probabilities given by the Born rule. Interpretations formulated after the Second World War, such as Bohmian mechanics and the Everett interpretation, would each in their own way avoid this rupture in the deterministic evolution of the quantum state, thereby offering two alternative ways of conceiving quantum processes without fundamental indeterminism. But this plurality of interpretations would become clearly visible only later. For several physicists of the interwar period, quantum mechanics therefore seemed to reopen a door that classical mechanics, its conservation laws, and its determinism had gradually closed on interactionist dualism.

Arthur Eddington was among the first to draw this consequence, in The Nature of the Physical World (1928). His position remained cautious: he proposed no precise mechanism by which mind could act on the brain, but argued that quantum indeterminacy removed a major metaphysical obstacle to human freedom. Since the world was no longer a rigidly determined block through and through, there was, in his view, no longer anything to prevent voluntary decision from finding a legitimate place in the openings left by physics itself. [Eddington (1928)]

Arthur Compton went further in his Terry Lectures, published as The Freedom of Man (1935). He defended the idea that mind might directly influence the outcome of indeterminate quantum events in the brain, and that this initially microscopic and indeterminate outcome could then be amplified into a determinate macroscopic effect—much as a tiny electrical signal can trigger a large-scale effect through an amplification mechanism. For Compton, such selection could allow the will to act on the body without necessarily injecting new energy into the physical system. The proposal nevertheless remained programmatic: it did not yet provide a mechanism showing how such selection could occur while respecting all the constraints of physical dynamics. [Compton (1935)]

This strategy would be taken up several decades later by John Eccles, who would try to locate, within the microscopic dynamics of the brain, a possible point of intervention for a non-physical mental cause. We will return to this attempt when discussing causal closure.

This attempt, however, failed to convince most physicists and philosophers of mind. The most immediate objection is that indeterminacy does not yet amount to action by the mind: replacing a strictly determined evolution with several probabilistic outcomes gives a non-physical will no control over which outcome actually occurs. One must still explain how the mind could select one outcome rather than another, and how that selection could fit into quantum dynamics without itself becoming an additional cause left outside the physical explanation.

Quantum indeterminacy therefore weakens one of the constraints inherited from classical mechanics—Laplacian determinism—but it refutes neither the conservation laws nor the more general principle of causal closure. The quantum interlude does not solve the interaction problem: it merely shifts the level at which a causal mechanism would have to be supplied, from the macroscopic to the quantum level.

VII. Physicalism Takes Shape: From Quantum Chemistry to Causal Closure

This section introduces the term physicalism, which should be distinguished from the materialism discussed so far. Classical materialism, as defended by La Mettrie, Diderot, or d'Holbach, holds that there is only one kind of substance: matter, generally conceived on the model of the physics of their time as extended, localized, and governed by mechanical laws.

The term physicalism emerged in the twentieth century, notably in the context of the Vienna Circle. This was a group of philosophers and scientists in the 1920s and 1930s who sought to construct a philosophy articulated as rigorously as possible with the sciences of their time. They were especially influenced by the revolutions of relativity and quantum mechanics, as well as by advances in logic and the foundations of mathematics. Their ambition was to bring philosophy and science closer together, clarify scientific language, and limit recourse to metaphysical speculation that could not be subjected to logical or empirical analysis.

In this context, however, the word “physicalism” did not yet mean exactly what it would come to mean in much of contemporary philosophy. For Otto Neurath in particular, it was first and foremost a thesis about language and the unity of the sciences rather than a metaphysical doctrine about the ultimate constitution of the world. The different sciences—physics, chemistry, biology, psychology, and sociology—should be able to communicate within a common, intersubjectively controllable language describing objects, events, and processes located in space and time. Physics provides the model and basic vocabulary for this project of unification, without this yet requiring the metaphysical claim that only physical entities exist. [Neurath (1931); Carnap (1932)]

The meaning of the term would later shift. In much contemporary philosophy, physicalism now primarily designates an ontological thesis: everything that exists is physical or depends entirely on the physical. The term thereby avoids some of the historical connotations of "materialism." Modern physics, after all, recognizes as real fields, spacetime structures, and other entities that no longer fit the classical image of small solid bodies in motion. Contemporary physicalism therefore does not claim that everything is material in the eighteenth-century sense, but that everything that exists belongs to the domain described by a complete physics, or depends entirely on that domain.

This reformulation makes the thesis more flexible in the face of changing physical theories, but it raises a difficulty: what exactly should "complete physics" mean? If it means current physics, physicalism risks being false because our theories are probably incomplete; if it refers to an ideal future physics, the thesis becomes harder to define without circularity. This is Hempel's dilemma, which I discuss at greater length in another article. [Hempel (1969), Vila Valls (2026)]

In contemporary usage, however, the terms materialism and physicalism are still often used as near-synonyms, depending on the historical context.

The explicit formulation of physicalism in the twentieth century did not arise solely from philosophical reflection on the mind. It also drew on the spectacular successes of physics when applied to domains that had previously been regarded as relatively autonomous, foremost among them chemistry.

In 1927, Walter Heitler and Fritz London proposed the first quantum treatment of the covalent bond in the hydrogen molecule6. Their model showed how the stability of a molecule could be understood through quantum mechanics and the interaction between electrons and nuclei. It thus became one of the founding achievements of quantum chemistry, which Linus Pauling and others would develop and systematize over the following years. For the first time, some fundamental chemical properties appeared explicable from the laws of the new physics rather than merely as empirical regularities peculiar to chemistry. [Heitler & London (1927)]

The exact significance of this success must nevertheless be specified. Heitler and London did not derive the whole of chemistry from the Schrödinger equation alone, and quantum chemistry does not provide, even today, a complete reduction of all chemical concepts and laws. Its calculations rely on approximations, models, and sometimes structural

[6] The hydrogen molecule, written H_2, consists of two hydrogen atoms, each composed of one proton and one electron. In a covalent bond, the two atoms share a pair of electrons. Chemists already used the idea of electron sharing before quantum mechanics, but there was not yet a satisfactory physical explanation of why this configuration produces a stable bond. Heitler and London's quantum treatment shows that the two electrons can form a common state in which the system's total energy is lower when the two nuclei are at a certain distance from one another than when they are very far apart. The bound molecule therefore corresponds to a more energetically stable state than the two separated atoms. Quantum mechanics thus provided, for the first time, a physical mechanism for understanding the stability of a covalent bond.

information that is not itself directly derived from fundamental laws alone. Concepts such as bond, valence, or molecular shape also retain an explanatory role of their own in chemical practice.

These limitations did not prevent the successes of quantum chemistry from producing, at the time, a profound impression of unification. In 1929, Paul Dirac could thus write that the fundamental physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry were now known, the main difficulty lying in the intractable complexity of the resulting equations. This statement should be understood as the particularly ambitious expression of a program: it claims that chemical phenomena share a common physical basis, not that all chemical explanations had already been replaced in practice by quantum derivations. [Dirac (1929)]

It was in this climate that the Vienna Circle—Carnap, Neurath, and Hempel in particular—developed its program of the unity of science during the 1930s: the various empirical sciences should be articulable within a common physicalist language. The advances of quantum chemistry provided this program with a particularly impressive model of success. They made plausible the idea that phenomena described at a higher level could depend entirely on physical mechanisms, without thereby demonstrating that the concepts and explanations of that level could be wholly eliminated or deduced from fundamental laws alone.[7]

It was in this context that mind-brain identity theory took its modern form, notably with Herbert Feigl in "The 'Mental' and the 'Physical'" (1958), and then with U. T. Place and J. J. C. Smart. These authors drew on an idea that had become increasingly plausible as the physical and biological sciences developed: the course of physical processes does not seem to require the intervention of any additional non-physical cause. In other words, the physical domain appears causally complete. [Place (1956); Feigl (1958); Smart (1959)]

This thesis, later called the principle of the causal closure of the physical, states that every physical effect has a sufficient physical cause. Feigl is among the authors who explicitly integrated it into the mind-body problem, but we should not attribute to him an isolated invention of the principle: it emerged instead from the progressive convergence of conservation laws, mechanistic physiology, neurology, and the growing unification of the sciences.

Causal closure then provides a powerful argument for mind-brain identity theory. If mental states genuinely produce physical behavior, while every physical event already has a sufficient physical cause, two possibilities seem to remain: either bodily effects are systematically produced by two distinct causal chains, one mental and one physical, or mental states are themselves identical with the physical states that play that causal role. Place, Feigl, and Smart regard the second hypothesis as more economical: pain, perception,

[7] On the distinction between physical dependence, reduction, and the explanatory autonomy of higher levels, see also: "Émergence, réduction et real patterns," The Naturalistic Outlook, https://thenaturalisticoutlook.com/2026/07/19/emergence-reduction-et-real-patterns/

or thought are not additional events accompanying brain activity, but that very activity described in another vocabulary.

In the same year, Paul Oppenheim and Hilary Putnam gave this movement a systematic formulation in "Unity of Science as a Working Hypothesis" (1958). They proposed a hierarchy of levels—from physical particles to atoms, molecules, cells, living organisms, and finally social groups—and envisaged that each level might, in principle, be connected to the level below through relations of micro-reduction. To make the idea completely explicit, phenomena characteristic of each level could, in principle, be explained in terms of the entities and processes of the immediately lower level: the properties of atoms in terms of the particles composing them; those of molecules in terms of atomic physics; cellular phenomena in terms of molecular mechanisms; organismic phenomena in terms of the functioning of cells; psychological phenomena in terms, among other things, of the activity of the nervous system; and finally social phenomena in terms of the behavior and interactions of individuals. The successes of atomic physics and quantum chemistry served as their model and strengthened the plausibility of this program. As the title of their article makes clear, however, this was explicitly a working hypothesis, not the claim that all reductions in the hierarchy had already been achieved. [Oppenheim & Putnam (1958)]

VIII. Philosophical Interlude: The Status of the Principle of Causal Closure

Before continuing the historical narrative, it is worth pausing over the exact nature of the principle of the causal closure of the physical, because its epistemic status is unusual and because it structures a large part of contemporary debates about mind.8

The first notable fact is that this principle is now widely accepted in naturalistic philosophy of mind, including by several authors who nevertheless reject reductive physicalism. It is not, however, the object of an absolute consensus: interactionist dualists and some emergentists continue to dispute it. Its influence comes above all from the pressure it exerts on any theory that attributes causal efficacy of its own to the mental. This pressure can be made transparent through the causal exclusion problem, formulated in its canonical version by the philosopher Jaegwon Kim in 1998.

The problem can be presented as a tension among several theses that, taken separately, appear plausible. The first is precisely the principle of causal closure, which states that every physical event has a sufficient physical cause. The second is the causal efficacy of the mental: we ordinarily attribute causal efficacy to our mental states; deciding to raise one's arm seems genuinely to contribute to the movement that follows. The third is the causal exclusion principle: except in special cases of genuine overdetermination, the same effect should not simultaneously have two wholly distinct sufficient causes. We generally attribute, for example, a particular cause to a person's death: illness, murder, a car accident,

---

[8]Epistemic here means "relating to knowledge and its justification." To ask about the epistemic status of a proposition is therefore to ask what reasons justify our holding it true and with what degree of certainty: logical demonstration, experimental result, inductive generalization, working hypothesis, and so on.

and so on. By extraordinary coincidence, someone might be struck by lightning at the very moment they are hit by a car, with either cause alone being sufficient to kill them instantly. That would be a case of overdetermination, but such cases are exceptional. We must distinguish this from cases in which the same causal sequence can be described at several levels. A general disease may cause a local dysfunction that leads to death: saying that the person died from the disease or from the dysfunction it caused does not necessarily amount to positing two independent causal chains. This is where the fourth thesis enters: the non-identity of the physical and the mental—the idea that physical brain processes and the mental states accompanying them constitute two distinct sets of facts. [Kim (1998)]

These four theses appear difficult to maintain simultaneously. For example, if one accepts causal closure, the causal efficacy of the mental, and the exclusion principle, it seems difficult to maintain the non-identity of the physical and the mental. If a bodily event such as shrugging my shoulders is caused both by physical processes in my brain and by the mental states accompanying them, and if there cannot be two different causal chains leading to the same effect, this seems to establish that mental states and physical states do not correspond to two distinct sets of facts. One is then pushed toward one form or another of physicalism. Still assuming causal closure, the two other solutions to the problem seem, at the very least, strange. The first would be to reject the causal efficacy of the mental: our thoughts would then not be responsible for the bodily movements we decide to perform. The second would be to reject the causal exclusion principle, which would mean that every voluntary movement is produced by a double causal chain. In this framework, the causal exclusion problem, together with acceptance of the causal closure of the physical, makes the physicalist solution extremely attractive.

The interactionist dualist can of course reject this conclusion by directly rejecting causal closure. But they must then specify where and how a non-physical mental cause intervenes in a physiological chain that increasingly appears describable in physical terms. John Eccles provides a particularly interesting case in this respect. A major neurophysiologist and yet an interactionist dualist, he did not merely assert in the abstract that mind acts on brain: he tried to identify the brain processes at which such action might occur.

In his later work, Eccles proposed that mental action could alter the probabilities of certain microscopic events associated with synaptic transmission, whose effects would then be amplified by neural-network activity. With Friedrich Beck, he notably developed a quantum model of synaptic exocytosis—the release of molecules at synapses that allows two neurons to communicate—in which a conscious intention could temporarily increase the probability of neurotransmitter release. Eccles also proposed describing certain functional units of the cortex as “dendrons,” corresponding to mental units called “psychons.” The appeal to quantum mechanics was intended precisely to make mental influence possible without supposing that an immaterial will simply injects energy into the brain. [Eccles (1992); Beck & Eccles (1992)]

The philosophical interest of this proposal is that it makes explicit what rejecting causal closure requires: it is no longer enough to assert that a mental cause might intervene; one must identify a point of intervention and propose a mechanism that could, at least in principle, be confronted with the study of the brain. Eccles's specific hypotheses have not, however, received recognized empirical confirmation. His case therefore illustrates both the logical possibility of rejecting closure and the empirical burden accompanying that rejection: as neural mechanisms are described without revealing any break that requires a non-physical cause, the interactionist must indicate with increasing precision where such a break should be sought.

The second notable fact is that causal closure is neither a logical theorem nor a law established by a single experiment. David Papineau presents it instead as an inductive generalization[9] gradually constructed through the history of science. Physical explanations have become increasingly successful at accounting for physical effects without the need to introduce additional vital, spiritual, or mental forces. The historical path traced so far—the development of conservation laws, the mechanization of physiology, advances in neurology, and the increasing physical integration of chemistry—is therefore not merely the intellectual background of the principle: it provides the core of the empirical argument in its favor. [Papineau (2001)]

For many naturalists, this generalization is all the more persuasive because it belongs to a broader history: the progressive replacement of explanations invoking intentions, purposes, vital forces, or spiritual principles by physical and physicochemical explanations. The movement begins with natural phenomena—thunder, lightning, earthquakes, or volcanic eruptions—then extends to living systems, through physiology, biochemistry, embryology, and heredity, before encompassing a growing share of mental processes. This history is also punctuated by phenomena once thought to lie, in principle, beyond the reach of natural explanation. Several of those boundaries have nevertheless receded. This obviously does not demonstrate that every remaining difficulty will one day receive a physicalist explanation; but for a naturalist it makes less persuasive the mere proclamation that some phenomenon must necessarily be inexplicable in physical terms.[10]

The history of celestial mechanics offers a famous example. Newton did not think that the order of the Solar System could be fully understood as the simple product of mechanical causes: he invoked God in his general explanation of the system's arrangement, and the mutual perturbations among the planets seemed capable of threatening its long-term stability. In the following century, Laplace showed that a significant part of these irregularities could instead be understood as consequences of the gravitational interactions themselves: some apparently cumulative perturbations are periodic and compensate for one another over very long timescales. It is in this context that we find the famous anecdote according to which, when Napoleon asked why God did not appear in his Mécanique céleste,

[9] An inductive generalization is an inference drawn from many convergent observations. If, for example, after centuries of studying tigers there has never been any report of a winged tiger, we can infer that tigers are animals without wings.

[10] An explanation is called teleological when it explains a phenomenon by the end or goal toward which it tends. Contemporary biology still commonly uses this kind of functional vocabulary, but generally seeks to explain it without invoking a conscious purpose or an independent vital principle, notably through natural selection and the regulatory mechanisms of organisms.

Laplace replied: “Sire, I had no need of that hypothesis.” The exact wording may be legendary, but it nicely summarizes the conceptual shift: what had seemed to require an external intervention became explicable through the laws of mechanics themselves. [Newton (1713); Laplace (1799–1825)]

Even the greatest physicists of the nineteenth century sometimes believed they had encountered such limits. In 1873, James Clerk Maxwell marveled at the extraordinary uniformity of molecules: molecules of the same type seemed to possess exactly the same properties wherever they were found in the Universe. Borrowing an expression from John Herschel, he argued that this identity gave them the character of “manufactured articles.” Maxwell took this to mark a limit to the explanatory power of natural science: in his view, neither the existence of molecules nor the exact identity of their properties could be attributed to any of the causes we call natural. He even declared that, on this point, “science must stop,” and concluded that the fundamental constituents of matter had been created. The problem he identified was real, but the limit he inferred from it was not: in the following century, quantum mechanics would show that the uniformity of atoms could itself receive a physical explanation, their properties following from the quantized structure of the states permitted by the laws of the theory. [Maxwell (1873)]

The same kind of proclamation can still be found today concerning open problems such as abiogenesis. We do not yet possess a complete and established account of the transition from prebiotic chemistry to the first living systems. But the present absence of such an explanation does not, by itself, imply that a non-physical principle is required. The preceding history instead encourages the naturalist to distinguish an unresolved scientific difficulty from an impossibility in principle.

This temptation to turn a present difficulty into a principled limit is a recurring theme. Gödel’s incompleteness theorem has, for example, been invoked by some authors—notably J. R. Lucas and later Roger Penrose—to argue that the human mind cannot be entirely computational[11]: for any sufficiently powerful formal system, there would be truths that it could not prove, whereas a human mind could, at least apparently, recognize their truth. This argument remains highly controversial, and Gödel’s incompleteness theorem is not, by itself, generally regarded as a refutation of computationalism. [Lucas (1961); Penrose (1994)]

Recent advances in artificial intelligence provide another example of this progressive naturalization of functions long associated with the mind. Artificial systems can now produce and interpret language, write programs, solve certain mathematical problems, or manipulate abstract concepts with an effectiveness that would have seemed spectacular a few decades ago. Yet there is no need to posit, in order to explain these performances, a cause lying outside the physical domain: they are realized by material architectures and computational processes wholly integrated within it. [Stanford HAI (2026)]

[11] To say that a mental process is computational is to say that it functions, at least in part, as information processing: some states serve as inputs, rules transform them, and other states result.

This obviously shows neither that the human mind is entirely computational nor that these systems are conscious. But it illustrates an important distinction: the fact that a capacity appears intellectually sophisticated, or even almost “magical” when observed, is not in itself a reason to think that it requires a non-physical mental principle. Artificial intelligence thus extends, onto new terrain, the historical movement by which functions once attributed globally to the mind become progressively accessible to mechanistic and physical explanation.

This justification of the principle of causal closure remains revisable. No finite series of observations can demonstrate that no non-physical cause will ever intervene in any process. The strength of the principle comes instead from the convergence of two observations: the persistent absence of recognized physical phenomena requiring a non-physical cause, and the growing ability of the sciences to complete their explanations without leaving the physical domain. Causal closure should therefore be understood as a general empirical thesis, strongly supported but in principle vulnerable to the discovery of a counterexample.

Several controversial research programs have nevertheless attempted to reveal phenomena that would appear incompatible with the causal completeness of the physical domain. Among the most frequently cited are Ganzfeld experiments, designed to test possible extrasensory transmission of information under conditions of reduced sensory input. Early series were criticized over randomization, possible sensory leakage, data selection, and statistical methods. Later improvements and automation of the protocols did not make the reported effects disappear entirely: several meta-analyses still report a mean result slightly above chance. The issue therefore remains genuinely controversial rather than simply refuted. These results have nevertheless failed to convince most scientists because the effect remains small, varies across studies and analytical methods, and has not produced a simple, sufficiently robust, independent replication establishing the existence of a transfer of information that cannot be physically explained. For my part, I regard this failure to achieve progressively greater clarity as an important reason for skepticism. Better protocols need not have increased the size of the effect, since early results may have been inflated by bias or inadequate controls; they should, however, have improved the signal-to-noise ratio, stabilized the estimate of the effect, and helped specify the conditions under which it appears. Yet after several decades of research, the Ganzfeld effect remains a small statistical excess whose size, robustness, and conditions of production are still disputed. This profile seems to me more compatible with a null effect accompanied by residual biases, methodological heterogeneity, and analytical flexibility than with the existence of a genuine channel of information transfer, without this conclusion amounting to a definitive logical refutation. [Storm et al. (2010); Hyman (2010); Tressoldi & Storm (2021)]

Near-death experiences provide another frequently cited argument. Some people who have been resuscitated report a sense of leaving the body, passing through a tunnel, seeing an intense light, encountering beings, or perceiving the medical scene from an external point of view. The subjective intensity and coherence of these experiences are not in dispute. But their main phenomenological components have analogues produced by hypoxia (lack of oxygen), certain psychoactive substances, anesthesia, epileptic seizures, brain stimulation, or other altered states of consciousness. More importantly, reports of allegedly veridical perceptions obtained outside the body generally rely on retrospective testimony whose exact timing and conditions of acquisition cannot be established with certainty. Prospective studies using information or targets placed outside the patients' field of view have so far produced no widely recognized case demonstrating the acquisition of information in the absence of brain function capable of supporting it. [Parnia et al. (2023)]

Neither Ganzfeld experiments nor near-death experiences can therefore simply be dismissed: they are legitimate objects of research, and some of the data remain debated. But they have not produced results robust, reproducible, and independent enough to be recognized as counterexamples to causal closure. Their present interest lies less in having refuted causal closure than in the possibility—always open in principle—of subjecting the thesis to empirical tests.

The principle of causal closure also operates, usually implicitly, in practices of reasoning outside the natural sciences. Forensic medicine and the law seek causal chains that can be established from natural mechanisms, material traces, and publicly checkable regularities. A court does not accept witchcraft, possession, or the intervention of an immaterial will as a competing explanation of an event simply because their absolute non-existence has not been demonstrated.

This does not mean that legal institutions officially adopt a physicalist metaphysical doctrine. Their commitment is primarily methodological and tied to standards of evidence: a merely imaginable possibility, unsupported by any positive evidence and not integrated into a reliable explanatory framework, does not constitute an admissible hypothesis. The presumption of innocence therefore does not require every conceivable supernatural cause to be treated as a reasonable doubt. In this practical and limited form, the idea that physical events should receive physically establishable explanations silently constrains the ordinary space of our causal reasoning.

One final technical clarification is needed. The formulation "every physical effect has a sufficient physical cause" seems natural within a deterministic physics: an earlier physical state and the laws then completely determine the later state. It becomes less clear when the fundamental theory assigns only probabilities to the different outcomes.

This does not mean, however, that indeterminism automatically opens a breach in causal closure. A probabilistic physical theory can be complete in the sense that it fully fixes the probabilities of the possible outcomes and leaves no explanatory role for an additional non-physical variable. A mental intervention selecting one of those outcomes would still constitute an addition to physical dynamics, even if it did not violate a strictly deterministic evolution.

It is therefore preferable to formulate the principle in terms of causal or explanatory completeness: for every physical event, physical laws and physical conditions provide all the causal explanation required by the best physical theory, whether they determine a unique outcome or only a probability distribution. This formulation remains neutral between fundamentally deterministic quantum theories, such as Bohmian and Everettian approaches, and theories introducing genuinely stochastic events, such as dynamical-collapse models. It preserves the core of the principle: no additional non-physical element is needed to complete the explanation of the evolution of the physical domain.

IX. From Behaviorism to Functionalism: From Disposition to Causal Role

We now return to our account of the rise of physicalism over the course of the twentieth century by turning to one of its most influential formulations: functionalism. This position can be understood as a point of convergence among several research programs. On the one hand, behaviorism had sought to define mental states scientifically through their relations to stimuli and observable behavior. On the other hand, cybernetics, the theory of computation, and the early computer sciences had shown that very different systems could be described in terms of the functions they perform and the relations among their inputs, internal states, and outputs, independently of the particular matter from which they are made. Functionalism would retain something of both intuitions while explicitly reintroducing internal states that behaviorism had largely eliminated.

John Watson's behaviorism, formulated in 1913, can be read as the culmination, on the side of behavioral science, of what Sherrington established for nerves and Loeb for organic matter: Watson refuses to invoke any mental cause in the explanation of behavior and wants psychology to deal only with directly observable and measurable stimulus-response relations. [Watson (1913)]

Behaviorism did not, however, remain confined to a methodological rule for experimental psychology. By the middle of the twentieth century it also received a philosophical formulation. Whereas Watson proposed that science should take as its objects only observable relations between stimuli and responses, logical behaviorism sought to analyze the very meaning of mental terms in terms of behavior or dispositions to behave.

To say that a person believes it is going to rain, is in pain, or wants a drink would therefore not refer to a private event taking place in an inaccessible mental substance. It would amount to saying that, in certain circumstances, the person is disposed to take an umbrella, complain, avoid injury, or seek water. This orientation appears in different forms in Carnap and Hempel, and later in Gilbert Ryle's The Concept of Mind, published in 1949. Ryle denounces as a category mistake the Cartesian idea that the mind is an additional thing, comparable to the body but hidden behind it—what he ironically calls the "ghost in the machine." He illustrates this kind of mistake with the example of a visitor who has been shown the colleges, libraries, lecture halls, and offices of a university and then asks: "But where is the University itself?" The question wrongly assumes that the University must be an additional thing placed alongside all the institutions that compose it. In the same way, according to Ryle, looking for "the mind" as a distinct entity located behind a person's behavior, dispositions, and capacities is to misunderstand what kind of thing the concept of mind refers to. [Ryle (1949)]

Philosophical behaviorism can thus dissolve part of the mind-body problem: if mental predicates refer to behavioral dispositions rather than to states of an inner substance, there are no longer two domains that must somehow interact. But this strategy seems to leave aside the existence of internal states that can vary without immediately producing observable behavior. A person may feel pain while concealing it; conversely, someone may simulate every behavior associated with pain without actually feeling it. These difficulties helped lead to the abandonment of strict behaviorism in favor of theories that reintroduced internal states—first by identifying them with brain states, then by defining them through their causal role within a broader organization.

Functionalism would emerge precisely from this double dissatisfaction: it retained the behaviorist idea that mental states should be characterized by what they do, while reintroducing internal states capable of playing a causal role between sensory inputs and behavior. Before functionalism fully developed, mind-brain identity theory provided a first way of bringing the mental back inside the organism by directly identifying mental states with brain states.

Mind-brain identity theory, introduced in Section VII, did not survive for long in its original form—and it is the revision it underwent that established the dominant theoretical framework for the following decades.

We must first specify what kind of identity is at issue. The theory defended by Place and Smart is generally understood as a type-type identity theory: each type of mental state would be identical with a particular type of brain state. The canonical example is pain identified with C-fiber activation. These fibers do indeed exist and participate in the transmission of certain signals associated with pain. But the equation "pain = C-fiber activation" should be understood here as a deliberately simplified philosophical illustration, not as an identification validated by neuroscience: conscious pain depends on much more complex processes, and the activity of these fibers is not sufficient, by itself, to explain it.

A type-type identity theory nevertheless asserts something of this form: wherever the same type of mental state occurs, the same physical type should be found. If pain were genuinely identical with C-fiber activation, no organism lacking such fibers could experience pain.

Hilary Putnam raised an objection to this thesis that became classic in "Psychological Predicates," published in 1967 and reprinted in 1975 under the title "The Nature of Mental States." This is the argument from multiple realizability. Very different organisms—human beings, other mammals, birds, or octopuses—seem capable of experiencing pain even though their nervous systems do not necessarily share the same physical organization. We can also conceive of a machine or an extraterrestrial built from a different substrate realizing the same mental state. If pain were strictly identical with one particular brain type, such systems could not feel pain unless they shared exactly that same physical type. [Putnam (1967)]

The idea can be understood through a simple analogy. The same kind of tool can be physically realized in very different ways. A mousetrap can operate by means of a spring, a trapdoor, a cage, or an electronic sensor. What allows these devices to be grouped into the same category is not the material from which they are made or the precise shape of their components, but the role they perform: responding to the presence of a mouse in such a way as to capture it.

For the functionalist, the same applies to mental states. Pain would not be defined by a particular nervous tissue, but by the place it occupies in the functioning of the organism: it is generally caused by injury or noxious stimulation, modifies beliefs and desires, draws attention, and tends to produce avoidance, withdrawal, or complaint. A human brain, an octopus nervous system, or an artificial device could therefore realize the same functional state through different physical mechanisms.

This thesis should not, however, be confused with token-token identity. A physicalist functionalist may maintain that every particular mental episode is realized by a particular physical event. But functionalism considered in isolation states first of all that mental types are defined by their causal roles; by itself, it does not settle the ultimate metaphysical status of each of their particular occurrences.

Putnam's appeal to functional vocabulary belongs to the broader context of the development of computing and cognitive science. The concept of a Turing machine[12] had shown that a computation could be defined abstractly by a set of states and transition rules, independently of the material from which the machine implementing it was made. The same computational organization can therefore be realized by different physical devices. [Turing (1936)]

Putnam transposed this model to the mind in what would be called machine functionalism. A mental state is defined by the place it occupies in the overall organization of the system: the inputs capable of producing it, its relations with other internal states, and the behavioral outputs to which it contributes. David Lewis extended this approach in 1972, in "Psychophysical and Theoretical Identifications," with a version known as causal-role functionalism: the mental terms of ordinary language—pain, belief, desire—implicitly define a causal role, and the corresponding mental state is simply whatever occupies that role in a given system, whether a human brain, an octopus brain, or, in principle, a silicon circuit. [Lewis (1972)]

Functionalism rapidly became the dominant position in philosophy of mind from the late 1960s through the 1970s, for a reason that went beyond philosophy alone: it offered a natural theoretical framework for the emerging cognitive sciences and for artificial intelligence. If the mind is defined by its functional organization rather than by its substrate, then in principle a suitably organized computer could think—a consequence Putnam explicitly accepted, and one that gave functionalism a close connection with the classical AI program of the period.

[12] A Turing machine is an abstract model of computation proposed by Alan Turing in 1936. Schematically, it consists of a device capable of reading and writing symbols on a tape and then changing state according to a set of rules. Its importance lies in showing that a computation can be defined by the organization of the operations performed, independently of the material substrate that implements them.

This victory did not go unchallenged. In “Troubles with Functionalism” (1978), Ned Block imagined a scenario that became famous: if the entire population of China, following a precise communication protocol, collectively simulated the functional organization of a human brain—each person playing the role of a neuron and transmitting signals according to the same causal relations—would this gigantic system constitute a genuine mind, with subjective experience? Block’s intuition is that it would not: the system would faithfully reproduce the required functional organization without anything guaranteeing that conscious experience resulted. This is the absent-qualia[13] argument: a system can, it seems, fully satisfy the functional description of a mental state without that state being accompanied by any lived experience. Faced with this objection, most functionalists choose to swallow the watermelon rather than abandon the thesis: they accept that, however strange it may seem intuitively, such a system would indeed satisfy the conditions required for subjective experience—the oddness of the conclusion not being, in their view, enough to invalidate the functionalist argument itself. [Block (1978)]

X. Toward the Hard Problem: From Nagel to Chalmers

The previous nine sections have told the story of a progressive victory of physicalism, consolidated by the causal closure of the physical. Yet it is precisely within this victorious camp, rather than in opposition to it from the outside, that a residual problem would re-emerge in the final third of the twentieth century—raised by authors who, for the most part, did not see themselves as opponents of physicalism, but as witnesses to its limits.

Thomas Nagel opens this sequence in 1974 with the famous article “What Is It Like to Be a Bat?” His argument begins from a simple observation: for every conscious organism, there is something it is like to be that organism—a subjective way in which the world appears to it. Nagel chooses the bat precisely because it perceives the world through echolocation, a sense humans lack: even if we knew absolutely everything about the bat’s physics, neurology, and behavior, nothing in that description seems to give us access to what it is like, from the bat’s own point of view, to perceive in that way. This subjective dimension of experience is what we call phenomenality: the fact that a conscious state has a certain lived character, a certain way of being experienced from the inside. Nagel’s point is methodological as much as metaphysical: physical science proceeds by objectification; it seeks to describe the world independently of any particular point of view—what is sometimes called a “view from nowhere.” Phenomenality, however, seems by its very nature tied to the existence of a subjective point of view. [Nagel (1974)]

It was Saul Kripke who, in Naming and Necessity (1980), unintentionally provided the conceptual tool that would allow his successors to turn Nagel’s intuition into a more explicit argument. Kripke begins from a principle: a true scientific identity discovered a posteriori—such as “heat is molecular motion”—is, if true, necessarily true; there is no possible world in which heat is not molecular motion. Put differently: before the nineteenth century, we did

[13] A quale (plural: qualia) is the qualitative character of a lived experience: what it is like, for example, to see red, feel pain, or smell an odor.

not know that heat corresponded to molecular motion. But once we understood that the causal consequences of heat could be fully derived from molecular motion, the connection between heat and molecular motion appeared necessary. Heat cannot be anything other than molecular motion, and molecular motion necessarily realizes heat. How, then, can we explain the fact that it nevertheless seems to us that things might have been otherwise? Kripke replies that this impression of contingency is a dissolvable illusion: what really seems possible is that molecular motion might have existed without producing in us the sensation of heat. Since "heat" designates the physical phenomenon itself rather than the sensation it usually causes, a gap can open between the two, and it is this gap that explains the illusion without threatening the identity. [Kripke (1980)]

Kripke then applies the same reasoning to pain and C-fiber activation. If this identity is true, it too must be necessary, and the impression that it is contingent should likewise be explainable as an illusion. But here the mechanism breaks down: unlike heat, there is no gap between pain and the sensation of pain—to be in pain just is to feel that sensation. We therefore cannot say that C-fiber activation might have existed without being felt as pain while nevertheless still being pain: if it is not felt as pain, then it simply is not pain. The gap that allowed the illusion to be dissolved in the case of heat is thus structurally absent here—not merely difficult to find for the time being, but impossible in principle, because pain is itself the sensation that the explanation is supposed to account for. The apparent contingency of "pain is C-fiber activation" therefore cannot be explained away as an illusion, which suggests—unlike in the case of heat—that it is not one: pain and C-fiber activation would then be genuinely distinct things.

Frank Jackson gives this kind of intuition its most memorable form with the knowledge argument in "Epiphenomenal Qualia" (1982). He asks us to imagine Mary, a brilliant neuroscientist who has spent her entire life confined to a black-and-white room but has had access to the totality of physical knowledge about color vision—every wavelength, every retinal process, every neural pathway involved. When Mary leaves her room and sees a red tomato for the first time, does she learn something new? Jackson's intuition, widely shared, is that she does: she learns what it is like to see red, information that no physical description, however complete, had given her. If Mary learns a new fact even though she already knew all the physical facts, then there must be facts—phenomenal facts—that do not reduce to physical facts. Jackson initially took this argument to support property dualism: brain states have physical properties (for example, neurons firing and communicating) and phenomenal properties (the famous qualia—for example, what it is like to see red or feel pain). It is noteworthy that he later reversed this conclusion: in "Mind and Illusion" (2003), he ultimately accepted a form of physicalism, arguing that the apparent novelty Mary experiences can be explained without positing non-physical facts—a rare reversal by the author of what is probably the most widely cited anti-physicalist argument of the twentieth century. [Jackson (1982, 2003)]

Between Kripke and Jackson on the one hand and Chalmers on the other, Joseph Levine introduced a decisive nuance in 1983 with the notion of the explanatory gap ("Materialism and Qualia: The Explanatory Gap"). Levine begins with a comparison: even if we accept that pain really is C-fiber activation—that is, even if we accept the identity—something is still missing that we possess in other cases of scientific identity. Knowing that water is H_2O allows chemical theory to explain why water boils at a certain temperature, why it is transparent, and why it dissolves salt: the microphysical structure explains the macroscopic properties. Nothing comparable exists for pain. Even if we accept that pain is C-fiber activation, no theory shows why that particular activation should produce that particular sensation rather than another one, or rather than nothing at all. Levine's move is nevertheless more cautious than Jackson's: he insists that this gap is epistemic—a deficiency in our present understanding—and that it does not necessarily imply an ontological conclusion. One can, in his view, remain a materialist while admitting that our current explanatory apparatus does not bridge the gap. [Levine (1983)]

It was David Chalmers who, in 1995 and then in The Conscious Mind (1996), brought these different difficulties together and systematized them. He first distinguished the "easy" problems of consciousness—explaining how the brain discriminates stimuli, integrates information, controls behavior, or enables a subject to report internal states—from the hard problem: why is the performance of all these functions accompanied by subjective experience rather than occurring with nothing being experienced from the inside? [Chalmers (1995, 1996)]

Chalmers then takes up and generalizes the modal strategy introduced by Kripke. Kripke had asked whether a particular brain state could exist without the experience with which it was supposedly identical. Chalmers extends this dissociation to the entire physical domain through the argument from philosophical zombies. A zombie would be a being physically and functionally identical to a human being—same brain, same behavior, same words, same reactions—but entirely devoid of subjective experience.

If such a being is metaphysically possible, then the complete set of physical facts is not sufficient to determine the phenomenal facts. Physicalism, understood as the thesis that a perfect duplicate of the physical must necessarily be a duplicate of everything that exists, would therefore be false. This conclusion does not, however, directly contradict the causal closure of the physical. Chalmers can accept that every physical effect has a complete physical explanation while maintaining that additional phenomenal properties accompany certain brain processes in virtue of psychophysical laws. His argument therefore targets the ontological reduction of consciousness to the physical, not the causal completeness of physical explanations.

As with Kripke, however, the entire difficulty lies in the transition from what we can conceive to what is genuinely possible. The mere fact that a zombie seems imaginable does not suffice to show that a world physically identical to ours but devoid of consciousness is metaphysically possible. [Chalmers (1996, 2002)]

Chalmers therefore tries to clarify what "conceive" means here. In particular, he distinguishes superficial conceivability, which may simply reflect our ignorance, from ideal conceivability: a situation would remain conceivable even after sufficiently thorough rational reflection and with all relevant information available.

This distinction matters because some scientific identities initially appear contingent even though they are in fact necessary. Before chemistry, one could, for example, imagine that water was not H_2O. But once it was discovered that the substance we call "water" is H_2O, there is no genuine possible world in which that same substance is something other than H_2O. What we had really imagined was a world containing a liquid resembling water but having a different chemical composition.

The "two-dimensional semantics" developed by Chalmers is designed precisely to formalize the difference between two questions: what we would have called "water" if the world had turned out to be different from what we thought, and what the substance actually designated by the word "water" could have been once its real identity was known. Chalmers argues that this strategy explains the apparent conceivability of the negation of many scientific identities—why we find it conceivable that water is not H_2O—without requiring us to conclude that there is a genuine difference. But, he argues, the same strategy does not dissolve the case of consciousness. According to him, even after all physical facts are fixed and all relevant information is available, a world physically identical to ours but devoid of subjective experience would remain conceivable.

This step remains one of the most disputed parts of his reasoning. His opponents can acknowledge a conceptual gap between physical and phenomenal descriptions while denying that this gap reveals a genuine metaphysical possibility. In particular, they argue that our phenomenal concepts may conceal a necessary identity with physical properties, so that our inability to deduce this identity—or to conceive the relation between brain and experience in any other way—would indicate a cognitive limitation rather than an ontological distinction. Chalmers thus goes further than Nagel, who remained agnostic, and Levine, who confined his gap to epistemology: he transforms a problem of description into a problem of ontology, and this is what made his formulation the standard reference point under the name "the hard problem of consciousness." [Loar (1990/1997)]

It is worth noting, at the close of this section, the structural kinship between this inference and the one that opened this article. Descartes' argument in Meditation VI—clearly and distinctly conceiving the mind apart from the body, inferring from this a genuine possibility, and then a distinction of substance—and Kripke's argument about pain and C-fibers, taken up and strengthened by Chalmers through philosophical zombies, share the same general pattern: from a failure to conceive things otherwise, or to dissolve a modal intuition, to an ontological conclusion about the distinction between things. The difference lies in the guarantee invoked. Descartes relied on a non-deceiving God to ensure that every clear and distinct conception corresponds to a genuine possibility—an external guarantee whose fragility in the face of Arnauld's objection was noted earlier in this article. Kripke, more cautiously, assumes no such guarantee: he systematically looks for a mechanism capable of explaining the appearance of contingency as an illusion, and concludes that there is a genuine distinction only where such a mechanism fails structurally. This caution does not, however, place the argument beyond an objection of the same general kind: the failure to explain away the illusion might reveal a limitation of our phenomenal concepts—opaque and insufficiently structured to reveal their physical realization conditions, as authors such as Brian Loar suggest—rather than a genuine metaphysical distinction. Three centuries separate Descartes from Chalmers, but both, in their own way, take the same step from what the mind cannot manage to conceive otherwise to what is really the case.

Conclusion

The history we have just traced first allows us to see how closely contemporary debates about consciousness are tied to successive transformations in the sciences. The mind-body problem has not remained unchanged since Descartes: it has shifted as physics, physiology, neurology, chemistry, and, more recently, the cognitive sciences have altered what it became reasonable to expect from a natural explanation.

Scientific progress has not, by itself, settled metaphysical questions. But it has profoundly transformed the space of available positions. Phenomena once explained through the action of agents, purposes, vital forces, or principles specific to living things have progressively been integrated into physical, chemical, and biological explanatory chains. The history of the problem of consciousness therefore illustrates a more general idea of philosophical naturalism: philosophical problems do not develop independently of the sciences. Scientific advances do not automatically dictate the correct metaphysics, but they continually change the constraints that metaphysics must satisfy, weaken some intuitions, and generate new ones.

It is in this context that the principle of the causal closure of the physical has acquired particular importance: the idea that the evolution of the physical domain requires no additional non-physical cause in order to be explained. This principle is neither a logical truth nor a universally accepted doctrine. It remains disputed, notably by some interactionist dualists and some emergentists. It nevertheless structures a large part of contemporary philosophy of mind because it makes especially difficult any theory that would attribute to a non-physical mind a causal efficacy of its own over the brain and behavior.

One essential point must therefore be emphasized: the hard problem of consciousness does not, in itself, challenge this causal closure. On the contrary, it arises largely within a framework in which closure is taken for granted. One can accept that bodily movements, perceptual discriminations, decisions, verbal reports, and the full range of functional processing carried out by the brain receive complete physical explanations while continuing to ask why these processes are accompanied by subjective experience.

In a sense, this is precisely what makes the problem especially difficult. As functions once attributed to a soul or a distinct mind become accessible to physical and functional explanation, phenomenality becomes increasingly isolated. The hard problem thus arises not despite the successes of naturalistic explanation but partly because of them: once the mechanisms of perception, behavior, memory, reasoning, and action control have been explained, there remains the question of why performing these functions is accompanied by something it is like to undergo them.

The historical path also reveals several of the major families of positions that we will encounter in contemporary debates. Materialism and then physicalism continue the strategy of treating mind as an aspect of the physical world itself. Dualism maintains, in deeply revised forms, that some dimension of mental reality cannot be identified with the physical. Neutral monism seeks a more fundamental reality of which the mental and the physical are two manifestations or aspects. Emergentism holds that certain properties can appear at higher levels of organization without simply being identical with the properties of their components. Mysterianism, finally, considers the possibility that the difficulty does not necessarily arise from the nature of consciousness itself, but from the cognitive limitations of the human mind when it tries to understand it.

These categories should not, however, create the illusion of a simple landscape. Each covers positions that can differ substantially, and some of the deepest disagreements divide authors who nevertheless belong to the same broad family. There are several physicalisms, several dualisms, several forms of emergentism and neutral monism, and their differences concern not only the status of phenomenality but also mental causation, reduction, explanation, and the relation between conceivability and metaphysical possibility.

The hard problem is therefore not an isolated question that suddenly appeared with Chalmers in the 1990s. Its contemporary formulation is the provisional outcome of a long transformation of the mind-body problem under the joint pressure of philosophical arguments and scientific discoveries. The second article will start from this point to examine the main responses proposed today: what they claim to explain, what they must abandon or revise, and what theoretical costs each is willing to pay in order to account for the existence of subjective experience.

References

Adrian, E. D. (1932). « The Activity of the Nerve Fibres ». Nobel Lecture, 12 décembre 1932.

Alexander, S. (1920). Space, Time, and Deity. 2 vol. London : Macmillan.

Arnauld, A. (1641). « Quatrièmes Objections », dans R. Descartes, Méditations métaphysiques.

Avery, O. T., MacLeod, C. M. & McCarty, M. (1944). « Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types ». Journal of Experimental Medicine, 79(2), p. 137-158.

Beck, F. & Eccles, J. C. (1992). « Quantum Aspects of Brain Activity and the Role of Consciousness ». Proceedings of the National Academy of Sciences, 89(23), p. 11357-11361.

Block, N. (1978). « Troubles with Functionalism », dans C. W. Savage (dir.), Perception and Cognition: Issues in the Foundations of Psychology, Minnesota Studies in the Philosophy of Science, vol. 9, p. 261-325. Minneapolis : University of Minnesota Press.

Born, M. (1926). « Zur Quantenmechanik der Stoßvorgänge ». Zeitschrift für Physik, 37, p. 863-867.

Broad, C. D. (1925). The Mind and Its Place in Nature. London : Routledge & Kegan Paul.

Buchner, E. (1897). « Alkoholische Gährung ohne Hefezellen ». Berichte der Deutschen Chemischen Gesellschaft, 30, p. 117-124.

Carnap, R. (1932). « Psychologie in physikalischer Sprache ». Erkenntnis, 3, p. 107-142.

Chalmers, D. J. (1995). « Facing Up to the Problem of Consciousness ». Journal of Consciousness Studies, 2(3), p. 200-219.

Chalmers, D. J. (1996). The Conscious Mind: In Search of a Fundamental Theory. New York : Oxford University Press.

Chalmers, D. J. (2002). « Does Conceivability Entail Possibility? », dans T. S. Gendler & J. Hawthorne (dir.), Conceivability and Possibility, p. 145-200. Oxford : Oxford University Press.

Compton, A. H. (1935). The Freedom of Man. New Haven : Yale University Press.

Crick, F. (1994). The Astonishing Hypothesis: The Scientific Search for the Soul. New York : Scribner.

Darwin, C. (1871). The Descent of Man, and Selection in Relation to Sex. London : John Murray.

Descartes, R. (1641). Meditationes de prima philosophia [Méditations métaphysiques]. Paris.

Diderot, D. (1769). Le Rêve de d'Alembert (rédigé en 1769 ; publié en 1830).

Driesch, H. (1908). The Science and Philosophy of the Organism: The Gifford Lectures Delivered before the University of Aberdeen in the Year 1907-1908. London : A. & C. Black.

du Bois-Reymond, E. (1872). Über die Grenzen des Naturerkennens. Leipzig : Veit & Comp.

Eccles, J. C. (1992). « Evolution of Consciousness ». Proceedings of the National Academy of Sciences, 89(16), p. 7320-7324.

Eddington, A. S. (1928). The Nature of the Physical World. Cambridge : Cambridge University Press.

Élisabeth de Bohême & Descartes, R. (1643). Correspondance sur l'union de l'âme et du corps, notamment les lettres de mai-juillet 1643.

Feigl, H. (1958). « The 'Mental' and the 'Physical' », dans H. Feigl, M. Scriven & G. Maxwell (dir.), Minnesota Studies in the Philosophy of Science, vol. 2, p. 370-497. Minneapolis : University of Minnesota Press.

Franklin, R. E. & Gosling, R. G. (1953). « Molecular Configuration in Sodium Thymonucleate ». Nature, 171, p. 740-741.

Heitler, W. & London, F. (1927). « Wechselwirkung neutraler Atome und homöopolare Bindung nach der Quantenmechanik ». Zeitschrift für Physik, 44, p. 455-472.

Hempel, C. G. (1969). « Reduction: Ontological and Linguistic Facets », dans S. Morgenbesser, P. Suppes & M. White (dir.), Philosophy, Science, and Method: Essays in Honor of Ernest Nagel, p. 179-199. New York : St. Martin's Press.

Hobbes, T. (1641). « Troisièmes Objections », dans R. Descartes, Méditations métaphysiques.

Hodgkin, A. L. & Huxley, A. F. (1952). « A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve ». The Journal of Physiology, 117(4), p. 500-544.

Holbach, P.-H. Thiry d' (1770). Système de la nature, ou Des lois du monde physique et du monde moral. Londres [Amsterdam].

Hubel, D. H. & Wiesel, T. N. (1962). « Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex ». The Journal of Physiology, 160(1), p. 106-154.

Huxley, T. H. (1874). « On the Hypothesis that Animals Are Automata, and Its History ». Nature, 10, p. 362-366.

Hyman, R. (2010). « Meta-Analysis that Conceals More than It Reveals: Comment on Storm et al. (2010) ». Psychological Bulletin, 136(4), p. 486-490.

Jackson, F. (1982). « Epiphenomenal Qualia ». The Philosophical Quarterly, 32(127), p. 127-136.

Jackson, F. (2003). « Mind and Illusion ». Royal Institute of Philosophy Supplement, 53, p. 251-271.

Jacob, F. & Monod, J. (1961). « Genetic Regulatory Mechanisms in the Synthesis of Proteins ». Journal of Molecular Biology, 3(3), p. 318-356.

Kim, J. (1998). Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation. Cambridge (MA) : MIT Press.

Kripke, S. A. (1980). Naming and Necessity. Cambridge (MA) : Harvard University Press.

La Mettrie, J. O. de (1747). L'Homme Machine. Leyde : Elie Luzac.

Laplace, P.-S. (1799-1825). Traité de mécanique céleste. 5 vol. Paris.

Leibniz, G. W. (1695). « Système nouveau de la nature et de la communication des substances ». Journal des savants.

Leibniz, G. W. (1714). Monadologie.

Levine, J. (1983). « Materialism and Qualia: The Explanatory Gap ». Pacific Philosophical Quarterly, 64, p. 354-361.

Lewes, G. H. (1875). Problems of Life and Mind, First Series, vol. 2. London : Trübner.

Lewis, D. (1972). « Psychophysical and Theoretical Identifications ». Australasian Journal of Philosophy, 50(3), p. 249-258.

Loar, B. (1990/1997). « Phenomenal States ». Philosophical Perspectives, 4 (1990), p. 81-108 ; version révisée dans N. Block, O. Flanagan & G. Güzeldere (dir.), The Nature of Consciousness (1997), p. 597-616. Cambridge (MA) : MIT Press.

Loeb, J. (1899). « On the Nature of the Process of Fertilization and the Artificial Production of Normal Larvae (Plutei) from the Unfertilized Eggs of the Sea Urchin ». American Journal of Physiology, 3, p. 135-138.

Loeb, J. (1912). The Mechanistic Conception of Life: Biological Essays. Chicago : University of Chicago Press.

Lucas, J. R. (1961). « Minds, Machines and Gödel ». Philosophy, 36(137), p. 112-127.

Malebranche, N. (1674-1675). De la recherche de la vérité. Paris.

Maxwell, J. C. (1873). « Molecules ». Nature, 8, p. 437-441.

McGinn, C. (1989). « Can We Solve the Mind-Body Problem? ». Mind, 98(391), p. 349-366.

Mill, J. S. (1843). A System of Logic, Ratiocinative and Inductive. London : John W. Parker.

Monod, J. (1970). Le Hasard et la Nécessité : essai sur la philosophie naturelle de la biologie moderne. Paris : Éditions du Seuil.

Morgan, C. L. (1923). Emergent Evolution. London : Williams & Norgate.

Nagel, T. (1974). « What Is It Like to Be a Bat? ». The Philosophical Review, 83(4), p. 435-450.

Neurath, O. (1931). « Physicalism: The Philosophy of the Viennese Circle ». The Monist, 41(4), p. 618-623.

Newton, I. (1713). « General Scholium », dans Philosophiae Naturalis Principia Mathematica, 2e éd. Cambridge.

Oppenheim, P. & Putnam, H. (1958). « Unity of Science as a Working Hypothesis », dans H. Feigl, M. Scriven & G. Maxwell (dir.), Minnesota Studies in the Philosophy of Science, vol. 2, p. 3-36. Minneapolis : University of Minnesota Press.

Papineau, D. (2001). « The Rise of Physicalism », dans C. Gillett & B. Loewer (dir.), Physicalism and Its Discontents. Cambridge : Cambridge University Press.

Parnia, S. et al. (2023). « AWAreness during REsuscitation – II: A Multi-Center Study of Consciousness and Awareness in Cardiac Arrest ». Resuscitation, 191, 109903.

Pauly, P. J. (1987). Controlling Life: Jacques Loeb and the Engineering Ideal in Biology. New York : Oxford University Press.

Penrose, R. (1994). Shadows of the Mind: A Search for the Missing Science of Consciousness. Oxford : Oxford University Press.

Place, U. T. (1956). « Is Consciousness a Brain Process? ». British Journal of Psychology, 47, p. 44-50.

Putnam, H. (1967). « Psychological Predicates », dans W. H. Capitan & D. D. Merrill (dir.), Art, Mind, and Religion, p. 37-48. Pittsburgh : University of Pittsburgh Press.

Ryle, G. (1949). The Concept of Mind. London : Hutchinson.

Sherrington, C. S. (1906). The Integrative Action of the Nervous System. New Haven : Yale University Press.

Sherrington, C. S. (1940). Man on His Nature. Cambridge : Cambridge University Press.

Smart, J. J. C. (1959). « Sensations and Brain Processes ». The Philosophical Review, 68(2), p. 141-156.

Sperry, R. W. (1968). « Hemisphere Deconnection and Unity in Conscious Awareness ». American Psychologist, 23(10), p. 723-733.

Spinoza, B. (1677). Ethica ordine geometrico demonstrata [Éthique].

Stanford Institute for Human-Centered AI (HAI) (2026). AI Index Report 2026. Stanford University.

Storm, L., Tressoldi, P. E. & Di Risio, L. (2010). « Meta-Analysis of Free-Response Studies, 1992-2008: Assessing the Noise Reduction Model in Parapsychology ». Psychological Bulletin, 136(4), p. 471-485.

Tressoldi, P. E. & Storm, L. (2021). « Stage 2 Registered Report: Anomalous Perception in a Ganzfeld Condition – A Meta-Analysis of More than 40 Years Investigation ». F1000Research.

Turing, A. M. (1936). « On Computable Numbers, with an Application to the Entscheidungsproblem ». Proceedings of the London Mathematical Society, 42, p. 230-265.

Vila Valls, A. (2026). « Émergence, réduction et real patterns ». The Naturalistic Outlook, 19 juillet 2026. https://thenaturalisticoutlook.com/2026/07/19/emergence-reduction-et-real-patterns/

Vila Valls, A. (2026). « Le naturalisme à l'épreuve du dilemme de Hempel ». The Naturalistic Outlook, 19 juillet 2026. https://thenaturalisticoutlook.com/2026/07/19/le-naturalisme-a-lepreuve-du-dilemme-de-hempel/

Watson, J. B. (1913). « Psychology as the Behaviorist Views It ». Psychological Review, 20, p. 158-177.

Watson, J. D. & Crick, F. H. C. (1953). « Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid ». Nature, 171, p. 737-738.

Wilkins, M. H. F., Stokes, A. R. & Wilson, H. R. (1953). « Molecular Structure of Deoxypentose Nucleic Acids ». Nature, 171, p. 738-740.