Philosophical Problems in the Study of Consciousness

Tobias Schlicht, *Ruhr-Universität Bochum* | 2025

Penultimate Draft | Comments welcome!

To appear in:
Olcese, U., Melloni, L. (2026) *The Scientific Study of Consciousness. Experimental and Theoretical Approaches*. Springer Nature.

Abstract

The scientific study of consciousness with its cornerstone of searching the neural correlates of consciousness constitutes an exciting and lively area of research. Yet, the empirical study of consciousness is surrounded by a host of philosophical challenges some of which are not new, while others have arisen and been elaborated over the last decades. The terminology can sometimes be confusing, causing misunderstandings how the challenges relate to each other. This paper provides an overview of several much-discussed philosophical problems of consciousness, offering a way of arranging them and some thoughts about how they relate. Most labels are taken from the literature so that there is a progression from the "easy" via the "hard" to the "harder" problem(s), and some labels are made up, like the "hardest" problem. For reasons of space, there is not detailed discussion of the possible solutions to these problems even though some problems have been formulated in response to others. For example, the "real" problem approach and the "illusion" or "meta" problems can be read partly as responses to the "hard" problem.

In 1994, Nobel laureate Francis Crick declared that "philosophers have had such a poor record over the last two thousand years" that they should leave the problem of consciousness to the experimental scientists, "learn enough about the brain to suggest ideas about how it works," and "learn how to abandon their pet theories when the scientific evidence goes against them" (Crick 1994, p. 258). Crick's book appeared with several other groundbreaking publications on consciousness, both from scientists (Blakemore & Greenfield 1987, Edelman 1989, 1992, Damasio 1994, Milner & Goodale 1995) and from philosophers (Churchland 1986, Baars, 1988, Dennett 1991, Flanagan 1992, Tye 1995, Dretske 1995, Chalmers 1996).[1]

Crick continued, "now is the time to think scientifically about consciousness (and its relationship, if any, to the hypothetical immortal soul) and, most important of all, the time to start the *experimental* study of consciousness in a serious and deliberate way" (Crick 1994, p. xii). President George Bush's declaration of the 1990s as the "decade of the brain" initiated a concerted empirical attack on consciousness, with the major research program of identifying the "neural correlates of consciousness". Apparently, it took the reputation of Nobel laureates like Francis Crick and Gerald Edelman to make consciousness a scientifically respectable topic of

[1] In addition, conferences on consciousness (e.g., TSC 1994 onwards), journals (e.g., Journal of consciousness Studies), and Associations (e.g., ASSC) were founded, leading to a massive increase in scientific publications on consciousness.

investigation.[2] The significant difference to earlier attempts at investigating consciousness scientifically in the 19th century was the availability of new methods and tools like fMRI to measure brain activity and allow investigations of consciousness from the outside rather than via introspection only (Revonsuo 2018, Frith & Rees 2007). Thirty years later, the science of consciousness is striving, a recent survey lists no less than 22 theories of consciousness that are or can be formulated in neurobiological terms (Seth & Bayne 2022) and a few of them are now being systematically tested and compared (Doerig et al. 2021, Lepauvre & Melloni 2021, Melloni et al. 2023, Cogitate Consortium et al. 2025). This is impressive but also a testament to the level of disagreement in the field.

This empirical research program was preceded by a long history of systematic philosophical discussions of consciousness, arguably initiated by Descartes' *Meditations* in the 17th century (Leach & Tartaglia 2016). But the new empirical approach raised many philosophical questions itself, rather than putting philosophy to rest, contrary to Crick's prognosis. In the last decades, an increasing number of different "problems of consciousness" have been formulated, with occasionally similar names which can easily lead to confusion. The goal of this review is to present several of these philosophical challenges and continuing debates, show how they relate to each other, and thereby provide a rough sketch of the current theoretical discussions, without aiming for completeness. The review starts with philosophical questions surrounding the correlations (section 1). The following sections introduce the popular distinction between the hard problem and the easy problems (section 2), the cluster of the illusion- and metaproblem (section 3) and the real problem (section 4) as responses to the hard problem. Section 5 presents the overflow problem, exemplifying a methodological challenge for the science of consciousness. Sections 6 and 7 discuss conceptual questions regarding the reach of phenomenal consciousness and its relation to self-consciousness. Section 8 turns to the so-called "harder" problem which concerns our epistemic situation vis-à-vis consciousness in radically different creatures and systems and illustrates an unresolved debate over functionalism and physicalism. The review closes with the hardest problem, which pays tribute to the foundational role of consciousness for science, which is considered the "blind spot" of cognitive science (section 9). Throughout the discussion, the philosophical problems are illustrated by focusing on particular theories as examples. For reasons of space, we can only hint at several ways in which philosophers and scientists are addressing (or trying to bypass) these challenges.[3] For the same reason, empirical evidence is discussed only anecdotally. Detailed evaluations of metaphysical options on the place of consciousness in nature have been left out (see the contribution by Antti Revonsuo to this volume).

1. Correlations and the Interpretation Problem

The research program proposed by Crick and his collaborator Christof Koch concerned the search for the neural correlates of consciousness (NCC), defined as "the minimal neural mechanisms jointly sufficient for any one specific conscious percept" (Koch 2012, p. 42). It was decidedly focused on neuroscience and *neural* mechanisms[4] and is still "arguably the

[2] See Damasio (1999, ch.1): "Studying consciousness was simply not the thing to do before you made tenure, and even after you did it was looked upon with suspicion. Only in recent years has consciousness become a somewhat safer topic of scientific inquiry."

[3] Accordingly, references in the text are taken to be exemplary to keep the text readable.

[4] Focusing on *minimal* mechanisms is supposed to rule out the (almost) trivial result that the whole brain is sufficient for consciousness – "almost" since one philosophical objection emphasizes the wider bodily and

cornerstone" of the scientific approach to consciousness (Chalmers 2000, p. 17). Since visual perception – whether conscious or unconscious – was the best understood sensory modality at the time, Crick and Koch (1990) started from evidence on vision to discover some neural process, activation pattern, property, brain circuit, or area that might be responsible for consciousness in *all* modalities.[5] Their favorite candidate was 40Hz oscillations which had been discovered to be involved in the *binding* of information about the same object (a flying round white ball, for example).[6] Information about distinct features of the object is processed in parallel and in a distributed fashion across the cortex. Wolf Singer and colleagues assumed that oscillations facilitated their integration into a unified percept (Gray and Singer 1989). Eventually, Revonsuo and Newman (1999) showed that binding via oscillations could happen unconsciously so that the role of such synchrony is not as straightforward as initially assumed. Many other candidate NCCs are discussed in Metzinger (2000).

Determining correlations between conscious and neural events faces empirical and methodological challenges and theoretical limitations (Aru et al. 2012). Since a correlation holds between two events, empirical evidence is compatible with several interpretations as to why the correlation holds. As Herbert Feigl recognized in his defense of the mind-brain identity theory, correlations merely support *parallelism*. Positing a *causal relation*, an *asymmetric dependence* (such as supervenience)[7] or even *identity* between the two events is not warranted by the data alone. Appealing to parsimony, he emphasized that "the step from parallelism to the identity view is essentially a matter of philosophical interpretation." (1958, p. 461) Apart from identity, several causal options are available. One of the events could have caused the other one, or they could both have been caused by a third, more fundamental event. Therefore, correlations are compatible with *idealism*, according to which the conscious event caused the neural activation, *neutral monism*, according to which a fundamental entity presents itself in two ways, as conscious experience and as a physical process, and with dualistic *epiphenomenalism*, according to which the conscious event is caused as a causally inert byproduct.[8]

Apart from parsimony, other considerations speak in favor of an identity theory. David Papineau (1998) put forward the causal argument for an identification of conscious properties with neural (i.e., broadly physical) properties. Since the causal efficacy of consciousness is worth preserving (Kim 1998) and since we assume that any *physical* event will have a chain of sufficient *physical* causes, the best way to preserve this efficacy is to identify the *conscious* causes with a subset of the *physical* causes, i.e., most likely neural events. Otherwise, we would have to accept either epiphenomenalism or a systematic causal overdetermination of many physical events. As Kim (2005) has shown, it is more difficult to preserve the causal efficacy of conscious properties if they are construed as irreducible non-physical properties. He eventually leans towards a view that had been championed before by David Lewis (1966, 1994) who used functional analyses of

---

environmental context in which brains are embedded and to which we turn below in section 5.1 (Noë & Thompson 2004).

[5] I will leave metaphysical questions aside and use the terminology of properties, events, processes and states loosely.

[6] Such oscillations and binding have also been discussed in the context of the unity of consciousness, a problem I won't discuss in this review (see Cleeremans 2003).

[7] For a discussion of Supervenience see Kim (1993) and Chalmers (1996).

[8] Searle's (1992) "Biological Naturalism" posits a causal relation from brain processes to conscious experiences, risking epiphenomenalism because consciousness remains irreducible and although he rejects the verbal dichotomy between 'mental' and 'physical', he cannot illuminate how consciousness can be both (see Kim 1998).

mental phenomena to yield species-specific identifications. Pain-in-humans would then be identical to some neural event, while pain-in-Martians or pain-in-snails would be identical to other physical events. This restricted type-identity, Lewis argued, would bypass the retreat to token identities typically advanced by functionalists who initially argued *against* type-identity theories. More on functionalism below.

Some philosophers think that the search for the NCC as defined above is doomed to fail. From the perspective of the enactive approach to cognition, the focus on neuroscience is misguided. Like trying to find the value of a Dollar bill in the texture of the paper, it is considered the wrong place to look for consciousness. Enactivists conceive of cognition and consciousness as activities carried out by embodied agents situated in an environment (Varela et al. 1991, Thompson 2007) rather than as brain states. They are informed and constrained by one's body (Gallagher 2005) and can extend into the agent's body and even into the agent's physical and social environment (Clark & Chalmers 1998, Vold 2015, Telakivi 2023). Thus, neither brain or nervous system nor skin constitute essential boundaries of a cognitive system. Consciousness, on this view, should be conceived not as an internal occurrence in the brain, "but rather as a complex set of capacities of embodied and situated agents" (Noë and Thompson 2004, p. 18-19).

Alva Noë and Evan Thompson (2004) criticize the "matching-content doctrine" (what David Chalmers (2000) calls a "content NCC"): For every experience, there is a neural representational system which is sufficient for the occurrence of the experience and whose content matches the content of the experience. Their first objection is that the active-attentional character of perceptual experience prohibits us from thinking of it as a kind of "snapshot". It is instead a "temporally extended process of attentive engagement" bringing forth perceptual content (Noë and Thompson 2004, p. 17):

> "After all, if perceptual content depends on the skillful activity of the whole animal or person, making use of its capacities for eye, head, and whole body movements, and for directed attention, then it becomes questionable whether there is any such thing as a minimal neural substrate sufficient to produce experience."

Their second objection is that there is nothing specifically visual or auditory about any neural activity such that NCCs could not explain the quality distinctive of individual experiences. Noë (2009) points to experimental evidence resulting from a rewiring of cortical connections in ferrets (Sur & Rubinstein 2005). Normally, information from their eyes ends up in the 'visual' cortex, processed via the LGN, information coming from their ears leads to the 'auditory' cortex, processed via the inferior colliculi. Removing the ferret's auditory pathway resulted in a new pathway that led from the eye to the auditory cortex, i.e., "the link between brain areas and conscious experience […] is malleable" (Noë 2009, p. 54). Since the rewired cortex was devoid of auditory input but computed the received visual input instead, this pattern of activity arguably resulted in the ferret's seeing (rather than hearing with their eyes).

Generalizing this point, O'Regan and Noë (2001) argued that the distinct qualitative character of seeing, touching, smelling, and hearing is due to the sensorimotor contingencies associated with any sensory modality. Our bodily constitution and sensorimotor capacities determine *what* we experience and *how*. We *hear* a noise (not just part of it) as coming from a certain direction, and while we can *grasp* only small objects like mandarins completely, we can *see* objects always

only in a certain profile, from a distinct spatial perspective. These contingencies differ across the senses and may inform (if not explain) the distinct sensory qualities. Whether the enactive approach has the tools to explain the distinct feel of conscious experiences (which we discuss in more depth in the next section) is questionable, but it points to an important limitation of the NCC research program and calls for a systematic clarification of how body and world contribute to the dynamics of consciousness. But while enactivists typically have a lot to say about the (neglected) contribution by body and world, they owe a detailed answer as to how we should interpret the brain's contribution to consciousness and mental phenomena more generally. They typically deny the mainstream account that it produces mental representations which eventually guide motor action, speech etc. Their positive account is often restricted and vague, e.g., "neural networks of perception are […] 'set up to be set off'" (Gallagher 2017, p. 115, Hutto and Myin 2017, p. 165), so that more work is needed here to replace the traditional story.

Further doubt on the possibility of finding NCCs comes from recent research on emotions. The classical picture of emotions assumes that emotions have unique fingerprints, constituted by special facial expressions (configurations of the 42 facial muscles for each emotion), significant bodily expressions manifested in heart rate or blood pressure, and special signatures in the brain. Lisa Feldman Barrett (2017a, 2017b) and colleagues scrutinized this line of work. Not only did it turn out to be impossible to identify a neural signature for fear, e.g., for which the amygdala had been a hot candidate. Considering the broader context of body functions and properties did not yield unique emotion patterns either. Rather than being the "seat" of fear, the amygdala showed increased activity to *any* face presented if the participant was previously unfamiliar with it, making novelty responsible for the effect (Weierich et al. 2010). The meta-analysis by Lindquist et al. (2012) found more evidence for the constructivist alternative, according to which many "interacting brain regions commonly involved in basic psychological operations of both an emotional and non-emotional nature are active during emotion experience and perception across a range of discrete emotion categories" (2012, p. 121). The *degeneracy effect* is that multiple combinations of neurons can yield the same result (Barrett 2017a). No populations of neurons or networks in the brain could be matched to any emotion category. Neither need instances of the same emotion category share a population of neurons for it to be implemented. Constructivists argue that "emotions should be modeled holistically, as whole brain-body phenomena in context" (Barrett 2017b, p. 16).

The lesson from failed attempts to identify neural signatures for emotions is that we should not expect to find them for consciousness. The claim that emotions, and mental phenomena in general, arise from the firing of neurons is compatible with the view that no single neuron is only involved in the making of an emotion. Drijvers, Small and Skipper (2025) have recently made an analogous case for the case of language, challenging the traditional view that the neural basis of language processing is localizable in a small set of brain regions. They claim that it is instead "widely distributed throughout the brain". Their exchange with Fedorenko et al. (2024, 2025) shows that the traditional view and reasoning behind it are far from commonly accepted. Analogously, any putative *minimal neural mechanism sufficient* for a conscious percept may be involved in unconscious processing in another context.[9]

[9] Pessoa (2022) questions the general assumption that a biological system like the brain can be neatly partitioned into functional units with clear demarcations and emphasizes that the brain must be investigated as a complex dynamical system. Regarding the role of the body, it has long been recognized that cephalopods display a unique distribution of neurons across their body (Godfrey-Smith 2018).

2. Reductive Explanation and the Hard Problem

The causal argument for the identity theory trades only on the possibility of mental causation and thus the *causal efficacy* of consciousness. It is silent on the *subjectivity* of consciousness. Appealing as the identity theory might be as the simplest explanation of a correlation, one is left wondering why some brain states do feel a certain way and others do not. David Chalmers (1996) introduced the notion of the "hard problem" for a science of consciousness to emphasize this sense of wonder. If physicalism were true, Chalmers argues, we should be able to understand why the neural basis of any phenomenal quality – be it seeing the blue sky or giving birth to a child – is the neural basis of *that* phenomenal quality rather than another one or none at all. Physicalism leaves an "explanatory gap" (Levine 1983, 2001) between brain process and experience.

The formulation of the hard problem has to do with the explanatory practice and resources available in cognitive science; it is contrasted with the so-called easy problems. Before we turn to them, a preliminary note on reductive explanation: Many philosophers of mind appeal to a form of reductive explanation characterized by a two-step procedure: It starts with providing a functional analysis of the phenomenon in question and then moves on to identify the physical mechanism responsible for carrying out the function. This is what David Lewis (1966) suggested, as already mentioned above. Here's a characterization of the relevant model by Jaegwon Kim:

> „To reduce a property, say being a gene, on this model, we must first 'functionalize' it; that is, we must define, or redefine, it in terms of the causal task the property is to perform. Thus, being a gene may be defined as being a mechanism that encodes and transmits genetic information. That is the first step. Next, we must find the 'realizers' of the functionally defined property – that is, properties in the reduction base domain that perform the specified causal task. It turns out that DNA molecules are the mechanisms that the task of coding and transmitting genetic information – at least, in terrestrial organisms. Third, we must have an explanatory theory that explains just how the realizers of the property being reduced manage to perform the causal task. In the case of the gene and the DNA molecules, presumably molecular biology is in charge of providing the desired explanations." (Kim 2005, p. 101)

For example, reproduction can be described as the "ability of an organism to produce another organism in a certain sort of way" (Chalmers 1996, p. 44). The second step involves identifying the mechanism that carries out this function in a range of systems. Thus, "once we have explained how those functions are performed, then we have explained the phenomena in question" (ibid.). Explaining reproduction requires only an identification of the mechanism allowing organisms to perform the function of producing another organism. In cognitive science, explaining learning requires only an identification of the mechanism responsible for performing the function of adapting "appropriately in response to environmental stimulation" (ibid.).

Chalmers (1995, pp. 200-201) lists a number of cognitive functions associated with consciousness that can be explained in this way, using the standard methods available to cognitive scientists: the ability to discriminate, categorize, and react to environmental stimuli, the integration of information, the ability to access and report mental states, the focus of attention, and the difference between wakefulness and sleep. Even if the *neural* mechanism cannot be identified, scientists can develop *computational models* capturing the functional task and setting

constraints on the putative realizing mechanisms. The functional characterization does not take brain processes to be exclusive but allows multiple possible realizers of the relevant function. Chalmers calls these phenomena *psychological* and the related challenges relatively "easy" because these phenomena can be exhaustively explained by appeal to structure and function. "Unfortunately," he argues,

> "[T]he kind of functional explanation that works so well for psychological states does not seem to work in explaining phenomenal states. The reason for this is straightforward. Whatever functional account of human cognition we give, there is a *further question*: Why is this kind of functioning accompanied by consciousness? No such further question arises for psychological states. […] Phenomenal states, unlike psychological states, are not defined by the causal roles that they play. It follows that explaining how some causal role is played is not sufficient to explain consciousness." (Chalmers 1996, p. 47)

Phenomenal consciousness is not characterized by its function but by how it *feels*, by how it *is* for a subject, and can consequently not be explained reductively in terms of structure and function. There is "something that it is like" (Nagel 1974) to see the blue sky or to have a conscious experience at all (see also Farrell 1950). Phenomenal consciousness is often illustrated with such examples as Frank Jackson's (1982) Mary, the color scientist, who learns everything physical there is to know about color vision but has never seen any colors. Does she learn something new when she finally sees the blue sky? Jackson argued that since Mary learns a new fact, not all facts are physical, and physicalism is false. After all, she knew all the physical facts before she saw colors and learned an additional, subjective fact. This is not the place to discuss all possible replies to this argument available to the physicalist, but a few possible avenues stick out. Lewis (1980) argued that Mary only acquires new know-how, but no propositional knowledge which leads – inter alia – to discussions about the relation between these kinds of knowledge (Stanley and Williamson 2001).

Another reply appeals to the so-called "phenomenal concept strategy" and holds that Mary learns a familiar fact in a new way since she acquires a new concept, a *phenomenal* concept, by *having* the color experience for the first time (Balog 1999). Rather than leaving the correlation of *two ontologically distinct properties* unexplained, the identity theory has the conceptual tools to explain how an *objective* property does not preclude *subjectivity.* Herbert Feigl had already suggested that there is only one event, a brain process, that can be referred to (or picked out) by two different descriptions or concepts. Just like the proper names "Norma Jean Baker" and "Marilyn Monroe" pick out one and the same person, the objective neurological concept "40 Hz oscillation" and the phenomenal concept "blue" could pick out the same subjective experience when seeing the blue sky, say. In this way, we can "substitute a dualism of concepts for a dualism of properties" (Block 2008, p. 401).

Physicalists can even concede the existence of subjective facts and reject the initial assumption of the argument that Mary can learn all the physical facts about color vision without ever having seen any colors. Robert Van Gulick notes that "if some physical facts are subjective (in the sense of being understandable only from a specific experiential perspective), then Mary could not even in principle know them given the stipulated restrictions on her past range of experience." (Van Gulick 2004a, p. 390) He distinguishes between the "physical science facts", which she could learn using books or a black-and-white screen, and "subjective physical facts", which are physically realized but "cognitively accessible only from the experiential perspective of a certain

range of physically realized self-understanding systems" (Van Gulick 2004a, p. 390-391). Since Mary had never been in the relevant brain state which is (according to the physicalist) identical to (or at least necessary for) the subjective experience of seeing the blue sky, she could not have had this knowledge. Therefore, the argument against physicalism would not get off the ground.[10]

This reply also applies to the example of wondering "what it is like" to be a bat (Nagel 1974) or trying to imagine what it would be like to have a child and become a parent, which includes "physically producing an infant, its immediate aftermath, and the extended experience of raising this child from infancy to adulthood" and pondering the responsibilities you are willing to take on and the sacrifices you are willing to make in becoming a parent (Paul 2014, p. 72). Facing this decision will involve a "transformative choice", as Laurie Paul calls it, since it is impossible to know what it is like to have a child until you have one, let alone give birth to a child before you do so. Paul argues that undergoing these experiences will be transformative because one could not have matched what it is like by trying to imagine these experiences and their consequences (feeling of responsibility etc.). You simply could not be in the relevant physical (brain) states realizing these experiences and consequently could not possess any knowledge about such subjective facts, whether they were physical or not.

Phenomenal realists believe that the subjectivity of consciousness poses an explanatory gap. Whether this *explanatory* gap yields an *ontological* gap, as Chalmers argues, is an additional question that we do not discuss here. It opens a spectrum of metaphysical alternatives to physicalism. Non-reductive options to be considered, apart from epiphenomenalism (the view that conscious properties are causally inert), are *panpsychism*, according to which some fundamental entities are conscious, and *panprotopsychism*, according to which these entities have properties that can produce consciousness in a larger system (Chalmers 2013). Another option, *neutral monism*, typically traced back to Ernst Mach and Bertrand Russell (see Stubenberg and Wishon 2023 for further references), could, Chalmers argues, take "information" to be fundamental. Typically, information is used in Claude Shannon's sense as observer-relative, but proponents of "Integrated Information Theory" (IIT) have coined an idiosyncratic notion of information that defines it as observer-independent (Tononi 2008). Since consciousness is taken to be observer-independent, information must likewise be so for them to be identifiable. IIT holds that the amount of integrated information can be measured in principle for any system and calculated mathematically. Since many systems in nature could then have a measure of integrated information greater than zero, IIT is often associated with panpsychism (Hassel-Mørch 2019).

Those who want to shy away from dualism or avoid such more extravagant metaphysical options may still accept that there is a hard problem, and an explanatory gap, but claim that it can be solved, and the gap can be closed. Further empirical or conceptual work was needed here, or so they would claim. It is not easy to evaluate whether and how accumulating ever more fine-grained empirical evidence about neural processing could solve this problem. More radical is the demand for a new "objective neuroscientific concept that would allow us to see how it could pick out the same phenomenon as our subjective concept" of the experience of blue, say (Block 2009, p. 1115). Thomas Nagel argues that our current phenomenal as well as neuroscientific concepts are inadequate since they are both limited to presenting one side of the correlation only. He claims that

[10] Cf. Ludlow, Nagasawa, and Stoljar (2004) for further discussion of the "Knowledge argument".

> "[T]he right point of view would be one that, contrary to present conceptual possibilities, included both subjectivity and spatiotemporal structure from the outset, all its descriptions implying both these things at once, so that it would describe inner states and their functional relations to behavior and to one another from the phenomenological inside and the physiological outside simultaneously—not in parallel." (Nagel 2004, p. 221)

Whether such concepts will ever be available is currently just as difficult to evaluate as the prognosis by eliminative materialists that when neuroscience will have matured, we will dispose of mental concepts and our folk-psychological taxonomy of mental phenomena, including its ontology of the mind, in favor of a thoroughly neuroscientific way of talking and explaining behavior (Churchland 1981).

3. Illusionism and the Meta Problem

A radical response to the hard problem and explanatory gap is the rejection of the explanandum phenomenal consciousness. On such a view, solving the easy problems, which is challenging enough, will not leave any aspect of consciousness unexplained. One version of this response is illusionism (Dennett 1991, Frankish 2016, Kammerer 2018, 2025). Illusionism was promoted by Daniel Dennett's (1988) critique of the philosophical notion of "qualia", often used to illustrate what is meant by phenomenal consciousness. He argued that the notion is incoherent since nothing could possess all their alleged features of being ineffable, intrinsic, private and infallibly known. Nowadays, the term "qualia" is much less popular than it was in the 1990s and early 2000s.[11] We do not need it for our purposes but will focus on phenomenal consciousness.

According to illusionism, we "never enter phenomenal states and there is nothing it is like to be anyone" (Kammerer 2019, p. 893). The key move is to argue that introspection is fallible and "delivers a partial, distorted view of our experiences, misrepresenting complex physical features as simple phenomenal ones" (Frankish 2016, p. 18). Or, as Kammerer (2019, p. 893) puts it: "We simply enter states (introspective states) that incorrectly represent that we are in phenomenal states, thus creating the illusion of phenomenal consciousness". The illusion is of a cognitive or introspective sort and thus different from perceptual illusions (e.g., the Müller-Lyer illusion). We are fooled into thinking that there is more to be explained than there is.

One central motivation for illusionism is that it provides a solid defense of physicalism without having to face the challenges that phenomenal realists must confront. Explaining the complex structure and function underlying consciousness, to use Chalmers' terminology, is sufficient to explain consciousness. Denying phenomenal consciousness helps avoiding the biggest obstacle for physicalism. This only holds for "strong illusionism", however, while any concession from "weak illusionists" who hold that phenomenal consciousness exists but does not have all the features typically ascribed to it, invites the hard problem through the back door (Kammerer 2019).

[11] Mathematical approaches to consciousness are sometimes formulated in terms of "qualia spaces" or "quality spaces" (Rosenthal 2015), attempting to capture the just noticeable differences between experiences formally. We will not engage with such approaches here, see Kleiner and Ludwig (2024) and Kleiner's contribution to this volume.

One major obstacle for illusionism is that its central claim is extraordinarily counterintuitive. After all, the alleged illusion is very peculiar and persistent and much more difficult to explain away than a perceptual illusion. Our task, according to the illusionist, is to "explain the content of the relevant states in broadly functional terms, and the challenge is to provide an account that explains how real and vivid phenomenal consciousness seems. This is the illusion problem" (Frankish 2016, p. 14). Put another way: "Illusionism replaces the hard problem with the illusion problem – the problem of explaining how the illusion of phenomenality arises and why it is so powerful." (ibid., p. 37)

The illusion problem aligns with what Chalmers (2018) has called the "meta-problem" of consciousness, which is so called because it is a problem about a problem. He defines it as the "problem of explaining why we think consciousness poses a hard problem", irrespective of whether one is a phenomenal realist (Chalmers 2018, p. 6). The data with which this problem is concerned are judgments and intuitions about consciousness (what Chalmers calls *problem reports*), which turn the meta problem into one of the easy problems, insofar judgements and intuitions are amenable to the explanatory methods of cognitive science.[12] He explicitly says that "to solve the meta-problem, [...] we need only explain the fact that we have the problem intuitions; we do not also need to explain their correctness." (Chalmers 2018, p. 21) The meta-problem must also be addressed by illusionists. Addressing the illusion problem is on a par with addressing the meta-problem. This is important, since a good response to these problems may persuade phenomenal realists to turn to illusionism. Chalmers at least believes a persuasive solution to the meta-problem can either help dissolve the hard problem or at least constrain the form of a solution.

Kammerer (2019, p. 850) introduces the "illusion meta-problem" as concerning a particular aspect of the illusion problem (Frankish) or meta problem (Chalmers), namely, its "peculiar strength". The task for the illusionist is to explain "why it is so hard for us, not only to *believe* that phenomenality is an illusion, but even to seriously *entertain* the possibility that it is". In contrast to perceptual illusions, we are not prepared to accept the categorization of conscious experience *tout court* as an illusion, we simply feel the urge to resist that move. That's why Galen Strawson calls it "the silliest claim" (2018, pp. 130-153). It sounds "crazy", as Kammerer admits. After all, one intuitive reaction to illusionism is to point out that in the case of consciousness we cannot create a chasm between how something *seems* and how something *really* is. It is the seeming that calls for explanation. This intuition supports the counterintuitive character of illusionism.

However, illusionists do not shy away from addressing the illusion meta-problem, although their account can only be mentioned here briefly. On Kammerer's version, phenomenal introspection depends on a subpersonal theory of mind, which provides phenomenal concepts as theoretical concepts. This cognitive mechanism which tracks our inner states misrepresents them as phenomenal experiences because the mechanism responsible for representing our phenomenal experiences is intimately intertwined with the mechanism by which we represent our epistemic situation. Consequently, "we grasp experiences as entities that cannot introspectively appear in

[12] Chalmers (1996) stipulates that judgements are characterized functionally. This is important because then our notorious Zombie Twins can make the same judgements as we do, despite the fact that, by Chalmers lights, our judgements contain phenomenal concepts which are partly constituted by experiences, while the Zombie's judgements contain concepts that must be different from ours. Lacking phenomenal experiences, Zombies cannot possess phenomenal concepts if these are supposed to be partly constituted by phenomenal experiences (Chalmers 2010).

a nonveridical way" (Kammerer 2019, p. 859). While things in the world can appear to us other than they really are, as is the case in perceptual illusions, experience is unique:

> „Any time we think about an appearance of an experience to a subject, we think about something that is a receptive affection which maximally resembles this experience, so that it has to *be* an experience of the same kind. For this reason, there cannot be a *nonveridical* appearance of an experience—this simply is a contradiction, given the content of our concepts." (Kammerer 2019, p. 861)

As becomes apparent, this account – although it can explain the missing appeal of illusionism – makes many contestable assumptions regarding introspection, its connection to the theory of mind, and the nature of phenomenal concepts so that there are ways in which the phenomenal realist can resist this explanation. Nevertheless, illusionism has become a popular position ever since proponents of various computational views of consciousness have combined their favored model of consciousness with a rejection of phenomenal consciousness. This can be illustrated with a brief look at Global Neuronal Workspace Theory (GNWT, this section) and the Predictive Processing Framework (next section).

For example, Stanislas Dehaene (2014, pp. 8-9, pp. 20-23; see also Dehaene & Naccache 2001) distinguishes three aspects that are crucial for the "transformation" of consciousness into a laboratory phenomenon, namely, vigilance (being intransitively conscious at all, ranging from full wakefulness to coma), attention, and cognitive access. What counts as

> "[G]enuine consciousness […] is conscious access —the simple fact that usually, whenever we are awake, whatever we decide to focus on may become conscious. Neither vigilance nor attention alone is sufficient. When we are fully awake and attentive, sometimes we can see an object and describe our perception to others, but sometimes we cannot—perhaps the object was too faint, or it was flashed too briefly to be visible. In the first case, we are said to enjoy conscious access, and in the second we are not" (Dehaene 2014, p. 9).

What makes Dehaene's approach align with Dennett's (1991) illusionist approach is his strong reliance on subjective reports (Dehaene et al. 2006, p. 206) and his rejection of phenomenal consciousness. Dehaene thinks that accepting this further explanandum

> "[L]eads down a slippery slope to dualism. We should start simple and first study conscious access. Once we clarify how any piece of sensory information can gain access to our mind and become reportable, then the insurmountable problem of our ineffable experiences will disappear." (Dehaene 2014, p. 10).

Note that Dehaene's approach requires us to "narrow our subject matter to a definite point that can be subjected to precise experiments" (2014, p. 8). Not all philosophers and scientists have to agree with this "preparation" of the explanandum. The quote above suggests that he does not expect any further step after "we start simple" with conscious access. The mystery will have disappeared, he believes. Believed, I should say, since he seems to have changed his mind and declared recently that the theory attempts to explain "conscious processing and "conscious phenomenology" (in Mudrik et al. 2025, p.2). While the shape of the theory has not changed, its target seems to have expanded to include phenomenal consciousness. We will return to GNWT and the methodological role of reports below in section 5.

4. Predictive Processing and the Real Problem

Anil Seth (2021) and Andy Clark (2024) defend the recently popular Predictive Processing (PP) framework which has been successfully applied to a wide range of perceptual and cognitive phenomena, but only recently to consciousness. PP gained significant popularity in the last decade partly because it promises to integrate perception, cognition, and action in one unified theory of mind. It is presented as aiming for explanations of all mental phenomena by postulating one distinct mechanism in the brain, namely Prediction Error Minimization (Frith 2007, Hohwy 2013, Clark 2016, Barrett 2017a, Seth 2021, Parr et al. 2022). In this framework, the brain is portrayed as a "sophisticated hypothesis tester" (Hohwy 2013, p. 2), constructing hypotheses about the distal causes of input on a hierarchy of computational levels, which are then matched against actual input to generate better predictions. The incoming neuronal signals are reinterpreted in this framework. Rather than counting as information needed for the brain to enrich a representation of the outer world, they count as error signals, transporting the divergences between predicted and actual sensory information. Depending on the estimated precision of these prediction error signals, the brain's model of the world is either amended and its current hypotheses are updated to accommodate the mismatch ('passive inference', or perception), or the hypotheses are held fixed and lead to resampling of the sensory states according to the current model ('active inference', or action).

Although PP is not a theory of consciousness, it is tentatively combined with illusionism by those who think that PP can help "supersede, dissolve, or deflate the 'hard problem' of consciousness" (Hohwy and Seth 2020, p. 21). More explicitly, Andy Clark endorses illusionism by forecasting that PP may "take the metaphysical sting out of the quale's tail" when "we put engineering, neuroscientific, and information-theoretic flesh on the familiar Dennett-style picture" (Clark 2019, p. 16). But illusionism is independent from PP. In Clark's approach to consciousness, illusionism rather than PP does all the heavy lifting (see Schlicht & Dolega 2021, Schlicht 2025 for further discussion).

Anil Seth (2021, p. 115) promises that "the hard problem intuition […] will fade away, eventually vanishing in a puff of metaphysical smoke", once we have built "explanatory bridges from the physical to the phenomenological". While Clark endorses illusionism explicitly, Hohwy and Seth don't and may not intend to subscribe to illusionism. They seem to suggest that further empirical work guided by the PP framework will eventually make headway on explaining phenomenal consciousness and make the hard problem go away. Anil Seth considers this task to constitute the "real problem": A science of consciousness should aim to "*explain, predict,* and *control* the phenomenological properties of conscious experience […] in terms of physical mechanisms and processes in the brain and body." (Seth 2021, p. 22) This sounds very similar to the canonical characterization of the hard problem. But according to Seth,

> „The real problem is distinct from the hard problem, because it is not – at least not in the first instance – about explaining why and how consciousness is part of the universe in the first place. It does not hunt for a special sauce that can magic consciousness from mere mechanism (or the other way around). It is also distinct from the easy problem(s), because it focuses on phenomenology rather than on function or behavior. It doesn't sweep the subjective aspects of consciousness away under the carpet. And because of its emphasis on mechanisms and processes, the real problem aligns naturally with a physicalist worldview on the relationship between matter and mind." (Seth 2021, p. 23)

In contrast to Clark's approach, Seth's real problem approach assumes that phenomenal consciousness is given and aims at building computational models that capture how neurophysiological and bodily processes are systematically linked to subjective experiences. It depends on available methodological means to access the "specific pattern of neural activity that is responsible for any and every experience, such as the experience of 'seeing red'. Whenever this activity is present, an experience of redness will happen, and whenever it isn't, it won't." (Seth 2021, p. 26) Seth is aware of the limitations of this approach, some of which have been mentioned above but advocates going beyond a science of correlations. This requires establishing explanations connecting properties of neural mechanisms with properties of subjective experiences that can illuminate why a certain brain activity should feel like seeing something red rather than something blue. He claims that the exact thing happened to our understanding of "life" which he takes as a source of optimism. Critics like Chalmers (1996, p. 108-109) object that the analogy does not hold since life can (in principle) be exhaustively characterized functionally, while consciousness cannot, which brings us back to the distinction between the easy and the hard problems, or between phenomenality and cognitive accessibility.

5. Methodology and the Overflow Problem

While David Chalmers distinguished between the hard and easy problems, Ned Block (1997) contrasted, in a similar vein, phenomenal consciousness with "access consciousness". A mental state is access-conscious if its content is available for the executive systems of reasoning and behavior, or more specifically, "if it is poised for free use in reasoning and for direct 'rational' control of action and speech" (Block 1997, p. 382). More recently, Block (2007) characterized access-consciousness not as a different *kind* of consciousness but renamed it "cognitive accessibility", as that what consciousness *does* in contrast to what it *is*, namely phenomenality.[13] GNWT (Dehaene 2014) can explain cognitive accessibility, Block maintains, but not phenomenal consciousness.

The disagreement has to do with the typical research methodology used by neuroscientists in their search for the NCC. Stanislas Dehaene and colleagues claim that "conscious perception must […] be evaluated by subjective report" (Dehaene et al. 2006, p. 206). Since he rejected phenomenal consciousness (see section 3), conscious access was the target of investigation. For a sensory representation to be consciously accessible and reportable, several factors are required: (1) vigilance, (2) sufficiently strong bottom-up activation in response to a stimulus, and (3) "the extension of brain activation to higher association cortices interconnected by long-distance connections and forming a reverberating neuronal assembly with distant perceptual

[13] In his earlier work, Block argued that phenomenal and access consciousness constitute different kinds of consciousness. To show that they can come apart he needed to present examples, but only discussed a putative one for phenomenal consciousness in the absence of access, e.g., when one has been consciously hearing a refrigerator humming or the noise from roadworks in the street outside but only becomes aware of it after a while. But he did not produce an empirical example for a case of access consciousness in the absence of phenomenal consciousness. He imagined a super-blindsight-patient, differing from an actual blindsight patient in their ability to spontaneously and voluntarily guess and report what – apparently – they do not see: "Visual information from his blind field simply pops into his thoughts in the way that solutions to problems [do]" (Block 1997, p. 385). Although the content of the relevant perceptual representation is poised for report and thus accessible, it is not phenomenal, according to Block, since there is still nothing it is like for them to see the stimuli. In his (2007), Block discussed the empirical data from Sperling (1960) to make a case for the richness of phenomenal consciousness in contrast to what can be cognitively accessed. But the claim that they constitute different *kinds* of consciousness does not play a major role anymore. The overflow argument turns on the richness of phenomenal *content*.

areas" (Dehaene et al. 2006, p. 205). Vigilance depends on the brainstem (*formatio reticularis* in particular) whose functioning is assumed to "awaken" the cortex and enables the cortical workspace. There is significant agreement among neuroscientists that consciousness *originates* in the brainstem (see Damasio 1999, Merker 2007, Solms 2021).

But on GNWT, vigilance is insufficient for conscious access which depends on "global information broadcasting within the cortex: it arises from a neuronal network whose raison d'être is the massive sharing or pertinent information throughout the brain" (Dehaene 2014, p. 13). Entering the workspace depends on the selection by (and respective bottleneck of) attention. Proponents of GNWT are impressed by experimental data on change blindness and inattentional blindness (Rensink et al. 1997, Simons & Chabris 1999) which allegedly demonstrate that we are not conscious of items outside attention. This is a contentious claim (Koch & Tsuchiya 2006, Block 2007). But the consequence is that, on GNWT, the NCC will involve the neural mechanism of attention and the mechanisms of report.

Moreover, the NCC also inevitably involves structures responsible for self-monitoring or metacognition, the monitoring of systemic processing, associated with prefrontal areas. Dehaene, Lau and Kouider (2017, p. 486) argue that the broadcasting of information via the global workspace and self-monitoring typically go together, although they can come apart when self-monitoring unfolds without being consciously reportable (e.g., automatic typing) or when reportable conscious processing is not accompanied by accurate monitoring (e.g., when some information suddenly pops up and creates a false memory). Nicholas Shea and Chris Frith (2019, p. 1) even argue that "the successful function of a global workspace critically requires that the broadcast representations include a metacognitive component" and thus implement metacognition, since the function of action guidance required that the representations in the global workspace be precise and carry confidence. Even if one rejects phenomenal consciousness, in a typical state of conscious access during wakefulness, the respective NCCs of vigilance, global broadcasting, and self-monitoring will all be activated, plus the neural mechanisms constituting the machinery of attention and report.

This confound gave rise to intense debate and further experiments. Block (2007, 2011) relied on different data to support his theory that attention does not restrict the capacity of consciousness but of cognitive accessibility such that the capacity of perceptual consciousness "overflows" the capacity of access. Experiments from Sperling (1960) and recent replications suggest that subjects can be conscious of much more information than they are able to cognitively access and report. These data contradict the prediction of GNWT that perceptual consciousness is sparse rather than rich.

When presented with an array of twelve letters (arranged in three rows), participants could report only four or five, but when Sperling only asked for a partial report, indicating the row to be reproduced with a tone after stimulus offset, participants produced the subset of letters without having known in advance which row would be required. Landman et al. (2003) revived this experiment with a change blindness paradigm, producing equivalent results. The information about the items had to persist and remain accessible after the presentation, but it is controversial whether all items had to be *consciously* perceived or not (Coltheart 1980). Block concludes that the "capacity" of phenomenal consciousness overflows "the working memory buffer that governs reporting" since subjects have "persisting experience as of more specific shapes" than can be brought under the specific concepts "required to report or compare those specific shapes with others" (Block 2007, p. 489).

According to GNWT, such neural representations are classified as “pre-conscious”, i.e., strong enough to enter the workspace but currently outside the focus of attention. According to Block (2007, p. 482-3), these data indicate that cognitive limitations of reporting could lead to phenomenally conscious experience one does not and cannot know about, as exemplified by a patient suffering from visuo-spatial extinction (Rees et al, 2000, Driver & Vuilleumier 2001).

The crucial question is this: When someone denies seeing object X, can we be sure that she does not phenomenally experience X? To decide this question, appeal to cognitive access and report is impossible, since regardless of whether the representation of X is (a) unconscious or (b) phenomenally conscious yet cognitively inaccessible, the subject will provide a negative answer. The first lesson to draw from this controversy is that the empirical evidence regarding the conceptual relation between attention, cognitive access, and phenomenal consciousness is pointing in two different directions. The second lesson is that to decide this issue empirically, the subject’s report must somehow be bypassed.

Victor Lamme (2006) proposed to *redefine* consciousness in neural terms, in terms of recurrent processing, using what he thinks to be the best candidate for the NCC. This would allow us to decide critical cases like blindsight, split brains etc., by measuring the presence or absence of recurrent processing, trumping whatever was reported. However, given the state of research into NCCs, this step would be premature. There are alternatives to recurrent processing being *the* NCC, and it would remain controversial whether recurrent processing in visual (or auditory etc.) areas alone was sufficient or whether global recurrent processing in the brain was required (as would be predicted by GNWT).

An alternative proposal appealed to “no-report paradigms” (Tsuchiya et al. 2015), one of which exploits the finding that certain automatic eye movements like the optokinetic nystagmus and the pupil reflex show a high correlation with conscious reports of perceptual dominance during binocular rivalry studies (Frässle et al. 2014). A comparison with a report paradigm yields a significant difference in the resulting NCC: the predominantly frontal activation was largely missing in the no-report paradigm, suggesting that it was associated with the participants’ introspecting, or accessing, their own perceptual experiences. Does this mean that no-report paradigms yield the NCC of pure experience without conflating it with cognition? Or is there a third way to decide this issue (Schlicht 2018)?

The overflow debate illustrates philosophical disagreement about the adequacy of the methods in determining the NCC, about the conceptual preliminaries informing this methodology, and the difficulty of turning “a philosophical mystery into a laboratory phenomenon” (Dehaene 2014, p. 8) while avoiding confounds. Moreover, it illustrates substantial disagreement among researchers about whether the content of consciousness is rich or sparse and about the role of attention for consciousness. Empirical researchers and philosophers have recognized this and expanded their agenda from searching for neural correlates to testing theories of consciousness, even in adversarial collaborative projects (Lepauvre and Melloni 2021, Melloni et al. 2023, Mudrik et al. 2025).

6. Understanding and the Range Problem

Phenomenal consciousness is typically associated with perceptual states, bodily sensations like pain, and affective states. But our daily conscious stream of consciousness involves a much

bigger range of mental phenomena. How many of them are or can be phenomenal? Is there something that it is like to have a conscious thought? Although many philosophers have answered this question negatively (Tye 1995, p. 4; Carruthers 2005, p. 138), others argue that there is, for example, something that it is like to *think that p* (Pitt 2004, 2024) or to *understand* something to be the case (Strawson 1994, p. 194), and that this phenomenology differs from, or goes beyond, sensory phenomenology. This is *cognitive phenomenology* (Bayne & Montague 2011).

A different discussion concerns the richness of the notion of phenomenal consciousness, not with respect to the range of mental states with qualitative character, or with respect to the richness of content that can be phenomenally experienced, as discussed in the preceding section. An additional aspect of richness concerns the question whether phenomenal consciousness involves reference to a subject of experience and whether this basic notion of consciousness therefore already constitutes a primitive or pre-reflective kind of self-consciousness (Van Gulick 2004b, Zahavi 2014, Schlicht 2017). We will sketch both discussions in turn (this section and section 7).

Chalmers' (1996) characterization of the hard problem may be partly responsible for the conservative position that phenomenal consciousness is restricted to the distinct qualitative characters of perceptions and sensations. Consider E.J. Lowe's early comment on Chalmers' presentation of the hard problem in which he objects to Chalmers' concession to functionalism and its underlying conception of cognition as information-processing:

> "Believing as he does that human thought and cognition in general are just a matter of 'information-processing', of a sort which could in principle go on in a mindless computer, he is left with the idea that all that is really distinctive about consciousness is its qualitative or phenomenal aspects (the 'what it is like', or 'inner feel'). And then it begins to look like a strange mystery or quirk of evolution that creatures like us should possess this sort of consciousness *in addition* to all our capacities for thought and understanding – these capacities being, for Chalmers, simply capacities for certain sorts of information-processing and storage. My response is that consciousness has only been put in this queer position by Chalmers (and, to be fair, by many others) *because* he has mistakenly denied it any role in his account of the nature of human thought and understanding. In short, it is the reductive, and wholly inadequate, *information-processing* conception of human cognition which is responsible for the misperception that 'consciousness' (in the form of 'qualia' and the like) occupies what threatens to be a merely epiphenomenal role as a peculiar additional feature of human mentation that is in no way essential to our basic intellectual capacities. [...] If [...] our capacity for genuine thought and understanding is quite inseparable from our capacity for phenomenal consciousness, then to the extent that Chalmers himself is correct in contending that reductive physicalism offers no prospect for an explanation of phenomenal consciousness, the conclusion ought to be that reductive physicalism, far from being equipped to solve the so-called 'easy' problems of consciousness, has in fact nothing very useful to say about *any* aspect of consciousness." (Lowe 1997, p. 121)

Rather than engaging with the conception of cognition as information processing, let's turn to considerations in favor of a richer notion of phenomenal consciousness. The position alluded to

in the quote above is that genuine thought and understanding are *inseparable* from, or even *grounded in*, phenomenal consciousness. This is sometimes called the "phenomenal intentionality research program" (Kriegel 2013). The main idea is that "real" or "genuine" intentionality is essentially phenomenal (Strawson 2004), while the intentionality of signs, public symbols, language, and photographs is derived from that "original" intentionality (Searle 1980).

This research program is opposed to the family of views under the heading "representationalism" which aim to explain phenomenal consciousness in terms (or as a version) of, mental representation (e.g., Tye 1995, Dretske 1995, Rosenthal 2005). First-order representationalists argue that a mental state is conscious if its non-conceptual content is poised for further cognitive processing, while higher-order representationalist views demand that a first-order state itself be represented by a distinct higher-order thought (Rosenthal 2005), guided by the idea that a representation is conscious if the subject is conscious *of* it (Brown 2025). Obviously, if genuine mental representation is grounded in, and thus presupposes, phenomenal consciousness, representationalism is doomed to fail. Angela Mendelovici (2018, p. xv), for example, defends the strongest version of this view, namely that "intentionality is simply identical to phenomenal consciousness". These two research programs are opposed to each other in several additional respects. For example, while representationalists typically subscribe to externalism about content, "phenomenal intentionality without compromise" is rigorously internalist in the sense that the content of thoughts depends only on facts internal to the subject (Farkas 2008).

The notion of cognitive, or "conceptual" phenomenology (Pitt 2024, p. 1) falls under this umbrella but is more constrained in that it need not apply to *all* instances of intentionality but may be restricted to thought or understanding. One important claim in this context is that there is a *distinct phenomenology of understanding* that is different from sensory phenomenology. Consider a woman who is given a white piece of paper with black ink marks on it. She bursts into tears. This reaction is only comprehensible on the assumption that the ink marks constituted a sad or terrifying message written in a language she understands. If there was merely sensory phenomenology, her experience of black on white, or even ink marks in a language she does not understand, could not explain her reaction and emotional state. Thus, the distinctive phenomenal quality of *understanding* is taken to explain why the woman reacts the way she does. Strawson (2004, p. 309) adds the qualification that "strictly speaking, it is the experience of *or as of* understanding these particular sentences, not the understanding itself […], for *mis*understanding the sentences would equally involve cognitive experience". Examples like this are supposed to pump the intuition that there can be, and is, something that it is like to think or understand that p. In his epistemological argument for this position, David Pitt (2024, p. 14) relies heavily on introspection which provides us – via direct acquaintance – with a "kind of first-person knowledge of the contents of conscious occurrent thoughts that we would not have if there were no such phenomenology". However, the epistemic credibility and adequate characterization of introspection has been subject to intense debate itself.

Follow-up questions concern the number of distinct types of phenomenology that there are and the way in which different sense modalities, thoughts and emotions are united or combined into the *single* unified experience we typically enjoy. Throughout the history of philosophy, the unity of consciousness (Bayne 2010) has both been treated as an important condition of conscious experience (for example, in Kant's *Critique of Pure Reason*) and as a fragile empirical datum (Nagel 1971). Kriegel (2015) distinguishes five irreducible, basic, sui generis "phenomenological primitives". Whether this is exhaustive and whether they are the right ones is open to debate. If

phenomenal consciousness plays a foundational or constitutive role for understanding and other cognitive states, then the hard problem becomes even more significant for cognitive science.

7. Subjectivity and egoless experience

Another line of argument to the effect that the notion of phenomenal consciousness captures more than merely qualia appeals to pre-reflective self-consciousness (e.g., Zahavi 2014). One possibility is to start by unpacking Nagel's famous phrase that there is "something it is like *for* the organism" (1974, p. 436). Consider Mary leaving her black and white room, having a conscious experience of the blue sky. There is a bluish way it is like for her to undergo that experience. Kriegel (2005, p. 23) suggests separating two aspects contained in this phrase, namely "(i) the *bluish* aspect, which we may call the experience's *qualitative character*, and (ii) the *for-me* aspect, which we may call its *subjective character*". While Mary's sensory experiences of seeing the blue sky, smelling roses, or listening to chirping birds *differ* in their qualitative character, they share the *same* abstract subjective character in that they present themselves as *hers*. The question is whether Mary's phenomenal experience is exhaustively characterized by adding the various qualitative characters of the elements making up her unified conscious experience or whether the subjective character makes up an additional aspect not captured in this summation.

As a corollary, a theory aiming to explain qualitative character may not be equally well equipped to explain subjective character. This may hold, for example, for the sensorimotor theory (O'Regan and Noë 2001), which appeals to the distinct sensorimotor contingencies associated with the various senses. While this may account for the different qualitative characters of visual, auditory, and other experiences, it may not provide a sufficient explanation for *subjective character.* Moreover, at least *conscious* thoughts may arguably contain subjective character, even if they lacked a distinctive phenomenology or qualitative character.

Unpacking Nagel's phrase in a different way, Robert Van Gulick (2004b) argues for a richer notion of phenomenal consciousness by appealing to self and world as constituting two sides of the same coin:

> "Experience is always the experience *of a self* and *of a world of objects.* It is both the experience *of a self located within a world of objects*, and also *of a world of objects as they appear and present themselves from the point of view of the self*. The phenomenal world and phenomenal objects are always apprehended from some self-like point of view, and conversely the phenomenal reality of such points of view consists largely in the access they afford for objects to appear or to present themselves within the experienced world." (Van Gulick 2004b, p. 81-82)

Finally, Zahavi (2014, p. 12), drawing on Sartre's (2003/1943) work, argues that the *for me-ness* or *mine-ness* associated with phenomenal consciousness constitutes the most basic kind of pre-reflective self-consciousness. After all, the distinct qualitative characters of seeing the blue sky, smelling roses, or listening to chirping birds, are not free-floating features, but experiential features presenting themselves *for* a subject of experience. There is something that it is like *for me* to enjoy ice-cream and there is something that it is like *for a bat* to be a bat.

One line of argument that supports resisting this intimate connection between phenomenal consciousness and self-consciousness appeals to, and explores, the notion of *pure consciousness* (Metzinger 2024). Methodologically, this research program relies heavily on empirical evidence from meditative practice and altered states of consciousness to test

> "[T]he working hypothesis […] that consciousness can exist not only in the absence of thought and sensory perception, but even without time experience, without self-location in a spatial frame of reference, and without any egoic form of bodily self-consciousness". In fact, the claim is that pure awareness "can be entirely dissociated from egoic self-awareness" and "exist without an experiential first-person perspective" (2024, p. xiii).

Obviously, if such pure experiences are possible, then Nagel's phrase and the above-mentioned interpretations do not capture all forms of conscious experience. One intuitive challenge for Metzinger's hypothesis is to explain how it is possible for subjects undergoing such pure conscious experiences to cognitively access and report them, if it is even rejected that the experiences presented themselves as *their own*. Arguably, if introspection can only *find*, but not *construct*, the mine-ness attaching to these experiences, it remains mysterious how they can be reported. Sascha Fink (2020, p. 1) presents it as a puzzle since if someone reports an alleged non-egoic experience, "the reporter (a) explicitly denies her existence during the selfless experience, but (b) implicitly affirms her existence as a witness to that selfless experience in order to give a first-person report about it". While Fink admits that "total ego-dissolution" is an incoherent notion, he argues that the relevant self-ascriptions can occur for a variety of reasons without causing contradiction (see also Costines et al. 2021).

The range problem and the problem of egoless experience are complex philosophical territories that require further exploration, and they build on a complex philosophical tradition. Different conceptions may have ramifications on debates of the hard problem, the distribution of phenomenal consciousness among biological creatures and artificial systems, and on the prospects of providing a reductive explanation of consciousness.

8. Other Minds and the Harder Problem

The range problem (section 6) leads to the distribution problem: Which creatures enjoy or possess consciousness? Put differently, which systems are included in the family for whom there is something that it is like to be them? Depending on how the criteria for consciousness are delineated, this family will be more or less inclusive. The problem of other *human* minds may be less mysterious and pressing than it is sometimes made to be. But patients with unresponsive wakefulness syndrome and patients in a minimally conscious state provide difficult cases where whether they are (still) conscious is important since much is at stake when doctors must make difficult decisions.

The question becomes even more puzzling once we enter the realm of other species. As Jonathan Birch (2024) has recently shown, the task of drawing a line anywhere in the animal kingdom or within the wider space of biological organisms is challenging enough. Once we consider neural organoids, xenobot technology (Blackiston et al. 2021), and artificial systems like Large Language Models or Social Robots, it becomes even more daunting. Are we limited to *attributing* consciousness only from the intentional stance (Dennett 1987) or can we decide cases on the 'edge of sentience'? Are there objective markers of consciousness that can be

applied across cases and species? Could we devise tests applicable to all candidate systems? (see Bayne et al. 2024 for discussion and references)

For systems whose constitution is very different from ours but still broadly biological, evolutionary considerations about the evolved mechanisms underlying consciousness can still be appealed to. But for artificial systems to be candidates for consciousness, the biochemical details of the implementation of consciousness in us (i.e., brains, neurons, neurotransmitters, etc.) must be negligible. The extent to which these details do matter determines the range of possible candidates. This is an empirical question. Functionalists and physicalists debate whether the appropriate causal-functional organization is crucial for consciousness, or whether details of the implementation of said functional organization are crucial as well. It is far from clear that we are in the epistemic position to decide which position is correct. Ned Block (2008, pp. 397-433) calls this the "harder problem" of consciousness.

To illustrate this, consider Global Neuronal Workspace Theory again. GNWT is presented as a computational theory: "consciousness is a computational property" (Mudrik et al. 2025, p. 2), identified as global broadcasting of information. GNWT is also typically portrayed as a version of functionalism. Dehaene (2014, p. 91) subscribes to functionalism but portrays it – somewhat vaguely – as the view that "consciousness is useful" (ibid., p. 91) and discusses it only in the context of fending off epiphenomenalism, i.e., the view that consciousness is causally inert. Functionalism is the view that, metaphysically, mental states are causal roles rather than biological states since these causal roles could be realized in different materials (Putnam 1965). Against the identity theory of mind and brain, popular in the 1950s, Putnam argued that this functionalist alternative was empirically more plausible.

Patrick Butlin et al. (2023) thus assume "computational functionalism" as a working hypothesis, since it allows for artificial consciousness *in principle*, and suggest using empirically supported neuroscientific theories to yield markers that can be used to assess consciousness in artificial systems. On the assumption that computational functionalism is true, this is a natural suggestion. Computational theories are functional, but functionalism and computationalism are independent. As Gualtiero Piccinini puts it:

> "It doesn't follow that if computationalism is true, functionalism is true; nor does it follow that if functionalism is true, computationalism is true. To get functionalism from computationalism, we also need an additional assumption, such as that the nature of mental states is (entirely) computational. To get computationalism from functionalism, we also need the independent assumption that all functional states are computational. Both of these assumptions are controversial; neither is especially plausible." (Piccinini 2009, p. 516)

Therefore, Butlin et al. are making two assumptions rather than one, and both must be defended against alternatives which emphasize the importance of the biochemical details of such markers. Proposing a computational function for consciousness does not turn the respective theory into a version of functionalism. It could be that even if mental states are in principle multiply realizable, once we take cognitive and conscious processes to be computations, all or some of them *may* depend for their implementation on highly constrained biological processes. The result will be a biological theory.

At this point, debates about the distribution of consciousness meet with debates about the nature of computation and its physical implementation (Anderson and Piccinini 2024). Proponents of computationalism typically emphasize that since computations are medium-independent, any system performing the relevant computation associated with consciousness will be conscious. According to Haugeland (1997, p. 10-11), "a concrete system is medium independent if what it is does not depend on what physical 'medium' it is made of or implemented in." Chess is medium-independent since its formal rules do not require a certain kind of board. We can play chess using only our imagination. Football or Snooker, by contrast, are medium dependent since they presuppose a pitch with certain features and a table of a certain size etc. While medium independence implies multiple realizability (associated with functionalism), the former is more stringent; multiple realizability does not imply medium independence.

When Haugeland introduced medium independence, he had digital computation in mind. However, it is questionable whether computations in the brain are digital. Maley (2021, 2023) argues that neural computation is analog and medium dependent, while Piccinini (2021) argues that neural computation is different from both digital and analog computation, constituting a computation sui generis. Anderson and Piccinini (2024, p. 254) take neural computation to be medium independent as well, rejecting any biochemical constraints on implementation. But if they take the brain's computations to be unlike any familiar computations we know, couldn't it be that their operation depends on the brain as their medium?

The crucial question for our purposes is whether the neural processing underlying consciousness, e.g., the putative global workspace and the function of broadcasting, are medium-independent. Cao (2022) and Chirimuuta (2022) argue it is not, or at least very unlikely. Putnam (1965) did not specify how it should be determined that functionalism is superior to a biological theory, but considering what we already know about communication across neurons, it is empirically more likely that it depends on biochemical features of the medium in which it takes place. After all, "inter-neuron communication mediated by spikes is not just electrical; it is chemical as well" (Chirimuuta 2022, p. 190). Pereda (2014) describes the differences between electrical and chemical synapses and their diverse forms of interaction, most of which will be medium-dependent to the same extent that chemical synapses depend on features of their chemical media. If such biochemical details are crucial for the "ignition" central for global broadcasting and the four neural signatures mentioned by Dehaene (2014), then GNWT may be empirically supported for consciousness in humans, but its computational version not so much. Dehaene holds, after all, that "there is *a system of neurons* that is able to select a piece of information, amplify it and broadcast it across the brain, and across modules" (in Mudrik et al. 2025, p. 2, emphasis mine). The point of the present argument is that we are not in an epistemic position to judge whether neural computations underlying consciousness are medium-independent. However, this would be needed to support computational functionalism. The computational version of a neuroscientific theory could simply yield a philosophical zombie devoid of phenomenality rather than a fully conscious system. Lacking access to the putative zombie's first-person data and absent neurophysiological markers it is unclear how one could determine the presence of consciousness in such a system. Block summarizes the problem like this:

> "Although functionalists are free to add restrictions, functionalism in its pure form is implementation independent. […] Functionalism and physicalism are incompatible doctrines, since a non-biological implementation of the functional

> organization of consciousness would be regarded as uncontroversially conscious by the functionalist but not by the physicalist. The big question for functionalists is this: 'How do you know that it is broadcasting in the global workspace that makes a representation conscious as opposed to something about the human biological realization of that broadcasting?' The problem for functionalists could be put like this: the specifically human biochemical realization of global availability may be necessary to consciousness—other realizations of global availability being 'ersatz'' realizations." (Block 2008, p. 119)

More recently, Anil Seth (2025) has made similar points, casting some doubt on the possibility of artificial consciousness, emphasizing that we should not even attempt to build such systems. While Dehaene, Lau and Kouider (2017) allow for consciousness in artificial systems *in principle*, they claim that currently available systems based on machine learning architectures operate unconsciously and do not meet the constraints set by neuroscientific theories. With regard to GNWT, it is an open question how much of the "messy biological details" (Cao 2022, p. 26) of neural computation can be ignored for global broadcasting to be implemented in a system with a radically different physical constitution and still yield consciousness. For an assessment of consciousness in artificial systems, metaphysical possibility à la Putnam is insufficient. Nomological possibility is required.[14]

Whether GNWT or other theories of consciousness are versions of functionalism or biologism is not only important for the question whether they allow for conscious *artificial* systems, but also for conscious *organisms* lacking the neuronal signatures associated with consciousness in primates like us. Many insights about consciousness have been obtained via experiments on monkeys and other animals (e.g., Logothetis & Schall 1989, Leopold & Logothetis 1996). Moreover, philosophical theories of consciousness have often been evaluated against the criterion whether they allow for consciousness in non-human animals. For example, the higher-order thought theory (Rosenthal 2005) faced the objection that its constraint of self-related thoughts was too sophisticated since it would deprive non-human animals and even infants of feeling pain. At one point, Peter Carruthers (1989, p. 509-511) bit the bullet and admitted this: His condition on a mental state's being phenomenally conscious was that it "is available to conscious thought". Since thoughts may require a natural language, he argued that "it seems highly implausible to ascribe such activities to any but the higher primates; and, even then, many of us would entertain serious doubts". Ever since, Carruthers has changed his mind on this point, and research on cognition and consciousness in animals and other organisms, including plants and slime mould, has been striving, with ramifications for ethical debates (Godfrey-Smith 2018, Andrews 2024, Birch 2024).

Phenomenal consciousness, as introduced by Nagel (1974), has typically been assumed to be widespread in nature and independent of sophisticated cognitive capacities: "Do animals other than humans undergo phenomenally conscious states? It certainly seems that way […] It would be absurd to suppose that there is nothing it is like for a dog that chews a favorite bone or a cat that prefers chopped liver for its dinner over anything else it is offered." (Tye 1995, p. 5) Peter Godfrey-Smith (2024, p. 10), tracing back the evolutionary origins of consciousness, assumes that "felt or conscious experience is probably widespread in animals […]". Philosophers pursuing

[14] Chirimuuta (forthcoming) makes this case for the prospects of neuromorphic computing. If the relevant functional equivalence between an artificial system and its biological model required data centers the size of Asia, the *in principle* possibility of artificial consciousness would not be very helpful.

the basal cognition research program go even beyond the animal kingdom and consider plants and other organisms as candidates for consciousness. Paco Calvo (2023, pp. 9-15), for example, is impressed by the fact that not only animals, but plants and even bacteria as well can be anaesthetized. Anaesthetized, a mimosa will no longer be able to fold its leaves, and the Venus flytrap's leaves will no longer snap shut:

> "The exciting – and controversial – implication is: if a plant can be temporarily put to sleep, as an animal can, does that mean it also has some kind of 'waking' state normally? Perhaps we might consider the possibility that plants are not simple automatons or inert, photosynthetic machines. We might begin to imagine that plants have some kind of individual experience of the world. They might be aware." (Calvo 2023, p. 15)

Among cognitive scientists and philosophers, such considerations are very controversial. But the point is not so much whether any side is right. The question is whether we will ever be in a position to find out, to measure or detect consciousness in such creatures from the outside. As Andrews et al. (2025) argue persuasively, various possible behavioral and neurophysiological markers for consciousness will differ across species and it may be difficult to decide which ones are decisive and how many must be present in each case. Jonathan Birch (2024, p. 17) puts it dramatically: "We start in a position of horrible, disorienting, apparently inescapable uncertainty about other minds, and then...the uncertainty is still there at the end. Sorry, it is inescapable. Anyone who tells you otherwise is not being honest or has not properly faced up to the problem." Birch argues for a "precautionary framework" which will ensure that in our attributions of consciousness to other systems, we "*err on the side of caution*". What's at risk is the possible suffering of (allegedly) impaired human patients, and of animals and other natural and artificial systems. In this vein, Andrews (2024) has argued for "shifting the null hypothesis" towards the *antecedent* assumption that "all animals are conscious", while Metzinger (2021) has called for a global moratorium on all research that might lead to consciousness in artificial systems at the risk of producing avoidable suffering.

There seems to be an asymmetry in the considerations about natural and artificial systems. While we are antecedently biased in assuming that non-human animals are likely candidates for phenomenal consciousness, such bias is lacking in the case of artificial systems. Given the assumptions about the wide distribution of phenomenal consciousness in nature, it is surprising that philosophers and AI researchers have considered only those artificial systems as candidates for phenomenal consciousness which are capable of sophisticated cognitive capacities, such as language production, abstraction, and possibly understanding (Chalmers 2023). Yet, faced with a Large Language Model, as impressive as its outputs may be, we are not antecedently convinced that it is a candidate system for phenomenal consciousness. By contrast, honeybees (Chittka 2022), for example, and other animals possess biological mechanisms with an evolutionary history that feeds into our antecedent conviction that they are candidates for consciousness. Jaan Aru et al. (2023) emphasize this and other points in their discussion of the prospects of conscious artificial systems from the perspective of neuroscience. They point to existing attempts to implement the GNWT architecture in artificial systems and take this approach to be feasible. But whether the perceiver architecture by Juliani et al. (2022) yields consciousness and whether and how this could be determined is precisely what is at issue here. As argued above, we cannot use AI systems to *test* theories of consciousness for their plausibility. We can only *apply* theories we are independently convinced of, to specific AI systems and stipulate that if the system meets the criteria for consciousness specified in the

theory, then we should judge the system to be conscious. However, it is unclear whether we presently have any such theories that we can be independently convinced of. The fundamental problem highlighted here is that we are not in the epistemic position to prefer a computational-functionalist theory over a neuroscientific one.

In his discussion of the harder problem, Ned Block discusses Commander Data from Star Trek who is superficially functionally equivalent to us but physically fundamentally different. It may well be possible that Commander Data was conscious and that this could someday become empirically testable. "But it is obvious that we do not now have any conception of how it could be tested" (Block 2008, p. 409) and "we have no conception of a ground of rational belief that Commander Data is or is not conscious, and we have no way of moving from a conclusion that Commander Data is conscious to any consequence for the truth of physicalism" (ibid., p. 416) The marker approach, as outlined by Bayne et al. (2024) and Passos-Ferreira (2024), seems promising for making progress on this question, at least regarding biological systems, but possibly also regarding artificial ones. Block's claim of the current inaccessibility of an answer to this epistemic question is the problem of other minds restricted to phenomenal consciousness and applied to the case of systems with a radically different physical constitution from ours, or of simply lacking a human brain (see McLaughlin 2003 for dissent). The fate of functionalism with respect to consciousness is thus far from settled, it cannot be simply assumed (see Seth 2024).

9. The Hardest Problem: The Blind Spot

The "hard problem" concerns the metaphysics of consciousness, the "harder problem" concerns an epistemic limitation which we are facing even if we had a solution to the hard problem for humans. But arguably the hardest problem concerns a systematic neglect of the fundamental role that consciousness plays for empirical discovery and theory building. In a recent book, Adam Frank, Marcelo Gleiser, and Evan Thompson (2024) call this problem the "blind spot". They argue that not only cognitive science, but all sciences, ignore that direct experience is "a precondition of observation, investigation, exploration, measurement, and justification" (Frank et al. 2024, p. xi).

Although the 'blind spot' is relevant for all sciences, cognitive science is special in the sense that it is "the science of the mind, and observation is a mental capacity based on conscious perception". Therefore, "when our concern is the mind, observation cannot be set aside or redefined operationally without presupposing what needs explaining" (Frank et al. 2024, p. 203). This prevents any attempts to explain consciousness reductively using methodologies that presuppose consciousness, i.e., in terms of entities or phenomena that are derivative to or acquired via consciousness:

> "The amnesia of experience eventually leads to a strange and nonsensical idea, prevalent in certain quarters of science and philosophy, that experience can be reduced to one or another of its structural residues. A striking case is the idea that conscious awareness can be reduced to the structural residue of informational or computational processes in the brain. This way of thinking inverts the whole procedure of producing objective knowledge by supposing that an abstract structural residue of experience (such as "information") can explain or ground the concrete being of conscious awareness. Here we encounter the Blind Spot at its most extreme." (Frank et al. 2024, p. 15)

As they recognize, most discussants ignore this problem and instead share the common ground of trust in the successful path of science, "scientific triumphalism" (Frank et al. 2024, p. ix), characterized by physicalism, reductionism, and objectivism. The mistake is to treat consciousness as an ordinary scientific topic by assuming that it can be explained as a natural phenomenon in terms of physical, or more specifically, neural mechanisms, using the explanatory tools of cognitive science. According to the blind spot argument, this is doomed to fail.

The Blind Spot argument can be illustrated by contrasting it with the hard problem. The point of Thomas Nagel's famous tale about what it is like to be a bat was to emphasize the uniqueness of consciousness compared to other objects of investigation because of the existence of a *subjective* perspective vis-à-vis the goal of science to attain *objectivity*. Nagel argued that in "other areas the process of reduction is a move in the direction of greater objectivity, toward a more accurate view of the real nature of things." (Nagel 1974, p. 444) But experience escapes this pattern, he continues, because "if the subjective character of experience is fully comprehensible only from one point of view, then any shift to greater objectivity – that is, less attachment to a specific viewpoint – does not take us nearer to the real nature of the phenomenon: it takes us farther away from it." (1974, p. 445) Nagel did not take this as sufficient reason to reject physicalism; he only concluded that we (at least currently) had no conception of how physicalism could be true. McGinn (1990) called this our "cognitive closure" to understand the truth of physicalism.

Frank, Gleiser, and Thompson make a more general point. Conscious experience is not confined to sounds, odors and the like as formulations of the hard problem may suggest. Rather, consciousness is what is presupposed by all scientific projects, and the blind spot of science is that human experience is the ineliminable root of science.[15] Consequently, "there is no way to step outside consciousness and measure it against something else. Everything we investigate, including consciousness and its relation to the brain, resides within the horizon of consciousness." (Frank et al. 2024, p. 186) Seen from this perspective, the hard problem is an artifact of the metaphysical picture associated with the "blind spot". It presupposes, as Thomas Nagel (2013), and long before him, Alfred North Whitehead (1920) have already pointed out, the "bifurcation of nature", isolating consciousness from the natural world, that began with the scientific revolution of the 17th century.

The typical move in the science of consciousness is to "surreptitiously substitute" any "abstract and idealized representation" generated based on discovery "for the concrete real world, the world that we perceive" (Frank et al. 2024, p. 8), ignoring that the representation presupposed direct experience. The authors discuss this mistake as they observe it in the Predictive Processing framework as well as in Integrated Information Theory. Mazviita Chirimuuta (2024) makes similar points regarding the computational theory of mind and the scientific realist attitude with which scientific findings and models are typically interpreted. We cannot discuss these points in depth here but it will be interesting to follow future discussions of this challenge to a science of consciousness.

---

[15] The blind spot metaphor refers to the presence of a blind spot in our visual field. The spot where the optic nerve enters the retina does not contain any light-sensitive cells. Since the spot is a natural consequence of the vision-enabling role of the retina and optic nerve, the missing information there has to be filled in. We don't experience the blind spot. Analogously, experience plays a science-enabling role but is not itself experienced as such – or occluded and forgotten (cf. amnesia) – in the process of science.

10. Conclusion

While consciousness has not been taken seriously as a respectable scientific topic before the 1990s, the science of consciousness is now striving, and it is exciting to follow the progress in this area. The goal of this paper was to provide an overview of some of the philosophical problems arising from the empirical research program of finding the neural correlates of consciousness in the brain. Empirical and theoretical progress in the last thirty years gave rise and was paralleled by a proliferation of problems, labeled as easy, hard, real, and harder, among others, which may lead to confusion once one enters the debate at any random entry point, or from different disciplines. Hopefully, this road map of discussions of philosophical problems and how they relate to each other, will be useful to philosophers and scientists alike. Needless to say, taking stock in this way will be superseded by further empirical work, so that, with some luck, some of the challenges for a science of consciousness can be overcome in the next decades.[16]

References

Anderson, N.G., Piccinini, G. 2024. *The Physical Signature of Computation.* Oxford: Oxford University Press.

Andrews, K. 2024. ‚All animals are conscious': Shifting the Null Hypothesis in consciousness Scienc. *Mind and Language* 39(3), 415-433.

Andrews, K., Birch, J., Sebo, J. 2025. Evaluating animal consciousness. *Science* 387, 822-824.

Aru, J., Larkum, M.E., Shine, J.M. 2023. The feasibility of artificial consciousness through the lens of neuroscience. *Trends in Cognitive Sciences* 46(12), 1008-1017.

Aru, J., Bachmann, T., Singer, W., Melloni, L. 2012. Distilling the neural correlates of consciousness. *Neuroscience and Biobehavioral Reviews* 36, 737-746.

Baars, B. 1988. *A Cognitive Theory of Consciousness*. Cambridge, Mass.: MIT Press.

Balog, K. 1999. Conceivability, possibility, and the mind-body problem. *Philosophical Review* 108, 497–528.

Barrett, L.F. 2017a. *How emotions are made. The secret life of the brain.* New York, Harper Collins.

Barrett, L.F. 2017b. The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience* 12(1), 1-23.

Bayne, T. 2010. *The Unity of Consciousness.* Oxford: Oxford University Press.

Bayne, T., Seth, A.K., Massimini, M., Shepherd, J., Cleeremans, A., Fleming, S.M., Malach, R., Mattingley, J.B., Menon, D.K., Owen, A.M., Peters, M.A.K., Razi, A., Mudrik, L. 2024. Tests for consciousness in humans and beyond. *Trends in Cognitive Sciences* 28(5), 454-466.

Bayne, T., Montague, M. (eds.) 2011. *Cognitive Phenomenology*. New York: Oxford University Press.

[16] I received valuable feedback from my team including Lara Bräuchle, Franziska Klasen, Bartosz Radomski, Caroline Stankozi, Tobias Starzak, and Elmarie Venter. I am also grateful for helpful comments from François Kammerer, Lucia Melloni and especially from Jaan Aru, and suggestions from an anonymous reviewer.

Birch, J. 2024. *The Edge of Sentience*. Oxford: Oxford University Press.

Blackiston, D., Lederer, E., Kriegman, S., Garnier, S., Bongard, J., Levin, M. 2021. A cellular platform for the development of synthetic living machines. *Science Robotics* 6, eabf1571.

Blakemore, C., Greenfield, S. (eds.) 1987. *Mindwaves: Thoughts on Intelligence, Identity, and Consciousness*. Oxford: Blackwell.

Block, N. 2011. Perceptual consciousness overflows cognitive access. *Trends in cognitive sciences*, 15(12), 567–575.

Block, N. 2009. Comparing the major theories of consciousness. In: M Gazzaniga (ed.), *The Cognitive Neurosciences* IV. Cambridge: MIT Press.

Block, N. 2008. *Consciousness, Function, and Representation. Collected Papers Vol. 1.* Cambridge, Mass.: MIT Press.

Block, N. 2007. Consciousness, Accessibility, and the mesh between psychology and neuroscience. *Behavioural and Brain Sciences* 30, pp. 481-548.

Block, N. 1997. On a confusion about a function of consciousness. In: *The nature of consciousness. Philosophical debates*. Ed. by N. Block et al. Cambridge: MIT Press, pp. 375-415.

Brown, R. 2025. *Consciousness as representing one's mind*. Oxford: Oxford University Press.

Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A, Deane, G., Fleming, S.M., Frith, C., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M.A.K., Schwitzgebel, E., Simon, J., VanRullen, R. 2023. Consciousness in Artificial Intelligence: Insights from the Science of Consciousness. *arXiv e-prints*. doi=10.48550/arXiv.2308.08708.

Calvo, P. 2023. *Planta Sapiens. Unmasking plant intelligence.* Boston: Little, Brown.

Cao, R. 2022. Multiple realizability and the spirit of functionalism. *Synthese* 200(506), 1-31.

Carruthers, P. 2005. Conscious experience versus conscious thought. In *Consciousness: Essays from a Higher-Order Perspective*. Oxford: Oxford University Press.

Carruthers, P. 1989. Brute Experience. *The Journal of Philosophy* LXXXVI, 63-79.

Chalmers, D.J. 2023. Could a Large Language Model be Conscious? *Boston Review* https://www.bostonreview.net/articles/could-a-large-language-model-be-conscious/.

Chalmers, D.J. 2018. The meta-problem of consciousness. *Journal of Consciousness Studies* 25 (9/10), 6-61.

Chalmers, D.J. 2013. Panpsychism and Panprotopsychism. https://consc.net/papers/panpsychism.pdf.

Chalmers, D.J. 2010. *The Character of Consciousness.* Oxford: Oxford University Press.

Chalmers, D.J. 2000. What is a neural correlate of consciousness? In: *Neural correlates of consciousness*. Ed. by T. Metzinger. Cambridge, Mass.: MIT Press.

Chalmers, D.J. 1996. *The Conscious Mind*. New York: Oxford University Press.

Chalmers, D.J. 1995. Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2(3), 200-219.

Chirimuuta, M. forthcoming. Neuromorphic computing and the significance of medium dependence.

Chirimuuta, M. 2022. The case for medium dependence: Comments on Neurocognitive Mechanisms by Gualtiero Piccinini. *Journal of Consciousness Studies* 29(7-8), 185-194.

Chirimuuta, M. 2024. *The Brain Abstracted*. Cambridge, Mass., MIT Press.

Chittka, L. 2022. *The Mind of a Bee.* Princeton: Princeton University Press.

Churchland, P. M. 1981. Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy* 78, 67–90.

Churchland, P.S. 1986. *Neurophilosophy*. *Toward a unified science of the mind/brain*. Cambridge, Mass.: MIT Press.

Clark, A. 2024. *The Experience Machine*. Oxford: Oxford University Press.

Clark, A. 2019. Consciousness as generative entanglement. *Journal of Philosophy* 116(12), 645-662.

Clark, A. 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.

Clark, A., Chalmers, D.J. 1998. The Extended Mind. *Analysis* 58(1), 7-19.

Cleeremans, A. (ed.) 2003. *The unity of consciousness*. *Binding, integration, and dissociation.* Oxford: Oxford University Press.

Cogitate Consortium, Ferrante, O., Gorska-Klimowska, U. *et al.* 2025. Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature* https://doi.org/10.1038/s41586-025-08888-1.

Coltheart, M. 1980. Iconic memory and visible persistence. *Perception and Psychophysics* 27(3), pp. 183–228.

Costines, C., Borghardt, T. L., & Wittmann, M. 2021. The Phenomenology of “Pure” Consciousness as Reported by an Experienced Meditator of the Tibetan Buddhist Karma Kagyu Tradition. Analysis of Interview Content Concerning Different Meditative States. *Philosophies* 6(2), Article 2. https://doi.org/10.3390/philosophies6020050.

Crick, F. 1994. *The astonishing hypothesis*. The scientific search for the soul. New York.

Crick, F. and Koch, C. 1990. Toward a neurobiological theory of consciousness. *Seminars in Neuroscience* 2, pp. 263-75.

Damasio, A.R. 1994. *Descartes’ Error. Emotion, Reason, and the Human Brain*. London: Penguin.

Damasio, A.R. 1999. *The feeling of what happens*. New York: Basic Books.

Dehaene, S. 2014. *Consciousness and the brain.* New York: Viking.

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., Sergent, C. 2006. Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences* 10(5), 204-211.

Dehaene, S., Lau, H., Kouider, S. 2017. What is consciousness and could machines have it? *Science* 358, 486-492.

Dehaene, S., Naccache, L. 2001. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79, 1-37.

Dennett, D.C. 1991. *Consciousness Explained.* Boston: Little.

Dennett, D.C. 1988. Quining Qualia. In: N. Block, O. Flanagan, G. Güzeldere, (eds.), *The Nature of Consciousness. Philosophical Debates.* Cambridge, Mass.: MIT Press, 619-642.

Dennett, D.C. 1987. *The Intentional Stance*. Cambridge, Mass.: MIT Press.

Doerig, A., Schurger, A., Herzog, M.H. 2021. Hard criteria for empirical theories of consciousness. *Cognitive Neuroscience* 12(2), 41-62, DOI: 10.1080/17588928.2020.1772214.

Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, Mass.: MIT Press.

Driver, J. & Vuilleumier, P. 2001. Perceptual awareness and its loss in unilateral neglect and extinction. *Cognition* 79(1–2), pp. 39–88.

Drijvers, L., Small, S.L., Skipper, J.I. 2025. Language is widely distributed throughout the brain. *Nature Reviews Neuroscience* brain. https://doi.org/10.1038/s41583-024-00903-0.

Edelman, G. 1989. *The Remembered Present. A Biological Theory of Consciousness.* New York.

Edelman, G. 1992. *Bright Air, Brilliant Fire*. *On the Matter of the Mind*. New York.

Farkas, K. 2008. Phenomenal intentionality without compromise. *The Monist* 91(2), 273-293.

Farrell, B.A. 1950. Experience. *Mind* 59(234), 170-198.

Fedorenko, E., Ivanova, A.A. & Regev, T.I. 2024. The language network as a natural kind within the broader landscape of the human brain. *Nat. Rev. Neurosci.* 25, 289–312. https://doi.org/10.1038/s41583-024-00802-4.

Fedorenko, E., Ivanova, A.A. & Regev, T.I. 2025. Reply to 'Language is widely distributed throughout the brain'. *Nat. Rev. Neurosci.* https://doi.org/10.1038/s41583-024-00904-z.

Feigl, H. 1958: The ›Mental‹ and the ›Physical‹. In: *Concepts, Theories, and the Mind-Body-Problem*. Minnesota Studies in the Philosophy of Science. Ed. by H. Feigl, M. Scriven and G. Maxwell. Vol. II. Minneapolis, 370–497.

Fink, S. B. 2020. Look who's talking! Varieties of ego-dissolution without paradox. *Philosophy and the Mind Sciences* 1(I/3), 1–36.

Flanagan, O. 1992. *Consciousness Reconsidered.* Cambridge, Mass.: MIT Press.

Frässle, S. et al. 2014. Binocular rivalry: frontal activity relates to introspection and action but not to perception. *Journal of Neuroscience* 34, 1738–1747.

Frank, A., Gleiser, M. and E. Thompson (2024). *The Blind Spot.* Cambridge, Mass., MIT Press.

Frankish, K. 2016. Illusionism as a theory of consciousness. *Journal of Consciousness Studies* 23(11-12), 11-39.

Frith, C. 2007. *Making up the Mind*. Oxford, Blackwell.

Frith, C., Rees, G. 2007. A brief history of the scientific approach to the study of consciousness. In: The Blackwell Companion to Consciousness. Ed. by M. Velmans and S. Schneider. Oxford: Blackwell.

Gallagher, S. 2017. *Enactivist Interventions.* Oxford: Oxford University Press.

Gallagher, S. 2005. *How the Body shapes the Mind*. Oxford: Oxford University Press.

Godfrey-Smith, P. 2024. *Living on Earth. Consciousness and the Making of the Natural World.* Glasgow: William Collins.

Godfrey-Smith, P. 2018. *Other Minds. The Octopus and the Evolution of Intelligent Life.* London: William Collins.

Gray, C.M., Singer, W. 1989. Stimulus-specific neuronal oscillation in orientation columns of cat visual cortex. *Proc. Nat. Acad. Sci* USA 86, 1698-1702.

Hassel-Mørch, H. 2019. Is the Integrated Information Theory of Consciousness compatible with Russelian Panpsychism? *Erkenntnis* 84(5), 1065-1085.

Haugeland, J. 1997. *Having Thought.* Cambridge, Mass.: Harvard University Press.

Hohwy, J., 2013. *The Predictive Mind*. Oxford: Oxford University Press.

Hohwy, J. and A. Seth 2020. Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences* 1(II), 12-30.

Hutto, D.D., Myin, E. 2017. *Evolving Enactivism.* Cambridge, Mass.: MIT Press.

Jackson, F. 1982. Epiphenomenal Qualia. In: *Philosophical Quarterly* 32, 127–136.

Juliani, A., Arulkumaran, K., Sasai, S., Kanai, R. 2022. On the link between conscious function and general intelligence in humans and machines. arXiv Published online July 19, https://doi.org/10.48550/arXiv.2204.05133.

Kammerer, F. forthcoming 2025. *House of Mirrors. The Illusion of Phenomenal Consciousness*. Oxford: Oxford University Press.

Kammerer, F. 2019. The illusion of conscious experience. *Synthese* 198(1), 845-866.

Kammerer, F. 2018. Can You Believe It? Illusionism and the Illusion Meta-Problem. *Philosophical Psychology* 31 (1), 44–67.

Kim, J. 2005. *Physicalism, or something near enough*. Princeton: Princeton University Press.

Kim, J. 1998. *Mental Causation and the Mind-Body Problem.* Cambridge, Mass.: MIT Press.

Kim, J. 1993. *Supervenience and Mind*. Cambridge: Cambridge University Press.

Kleiner, J., Ludwig, T. 2024. What is a mathematical structure of consciousness? *Synthese* 203: 89, doi.org/10.1007/s11229-024-04503-4.

Koch, C. 2012. *Consciousness. Confessions of a romantic reductionist*. Cambridge, Mass.: MIT Press.

Koch, C., Tsuchiya, N. 2006. Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences* 11(1), 16-22.

Kriegel, U. 2015. *The Varieties of Consciousness.* Oxford: Oxford University Press.

Kriegel, U. (ed.) 2013. *Phenomenal Intentionality.* Oxford: Oxford University Press.

Kriegel, U. 2005. Naturalizing subjective character. *Philosophy and Phenomenological Research* LXXI, pp. 23-57.

Lamme, V. 2006. Towards a true neural stance on consciousness. *Trends in Cognitive Sciences* 10(11), 494–501.

Landman, R., Spekreijse, H., Lamme, V.A.F. 2003. Large capacity storage of integrated objects before change blindness. *Vision Research* 43, 149–164.

Leach, S., Tartaglia, J. (eds.) 2016. *Consciousness and the Great Philosophers. What would they have said about our mind-body problem?* London: Routledge.

Leopold, D. & Logothetis, N.K. 1996. Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379, 549–553.

Lepauvre, A., Melloni, L. 2021. The search for the neural correlate of consciousness. Progress and challenges. *Philosophy and the Mind Sciences* 2(4).

Levine, J. 2001. *Purple Haze. The Puzzle of Consciousness.* Cambridge, MA.

Levine, J. 1983. Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly* 64, 354–361.

Lewis, D.K. 1980. What experience teaches. In: W.G. Lycan (ed.), *Mind and cognition: a reader*. Cambridge, Mass.: Blackwell, 29-57.

Lewis, D.K. 1994. Reduction of Mind. In: *A Companion to the Philosophy of Mind*. Ed. by S. Guttenplan. Cambridge: Blackwell, pp. 412-431.

Lewis, D.K. 1966. An argument for the identity theory. *Journal of Philosophy* 63(1), 17-25.

Lindquist, K.A., Wager, T., Kober, H., Bliss-Moreau, E., Barrett, L.F. 2012. The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences* 35, 121-202.

Logothetis, N. K. & Schall, J. D. 1989. Neuronal correlates of subjective visual perception. *Science* 245, 761–763.

Lowe, E.J. 1997. There are no easy problems of consciousness. In: J. Shear (ed.): *Explaining consciousness. The hard problem.* Cambridge, MA.

Ludlow, P., Nagasawa, Y., Stoljar, D. (Eds.) 2004. *There's something about Mary. Essays on Phenomenal Consciousness and the Knowledge Argument.* Cambridge, Mass.: MIT Press.

Maley, C. 2023. Medium independence and the failure of the mechanistic account of computation. *Ergo* 10, 28. https://doi.org/10.3998/ergo.4658.

Maley, C. 2021. The physicality of representation. *Synthese* 199, 14725–14750.

McGinn, C. 1990. Can we solve the mind-body problem? *Mind* 98, 349–366.

McLaughlin, B. 2003. A naturalist-phenomenal realist response to Block's harder problem. *Philosophical Issues* 13, 163–204.

Melloni, L., Mudrik, L., Pitts, M., Bendtz, K., Ferrante, O., Gorska, U., Hirschhorn, R., Khalaf, A., Kozma, C., Lepauvre, A., Liu, L., Mazumder, D., Richter, D., Zhou, H., Blumenfeld, H., Boly, M., Chalmers, D.J., Devore, S., Fallon, F., de Lange, F.P., Jensen, O., Kreiman, G., Luo, H., Panagiotaropoulos, T.I., Dehaene, S., Koch, C., Tononi, G. 2023. An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory. *PLoS One*. 18(2):e0268577. doi: 10.1371/journal.pone.0268577.

Mendelovici, A. 2018. *The Phenomenal Basis of Intentionality.* Oxford: Oxford University Press.

Merker, B. 2007. Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences* 30(1), 63-81.

Metzinger, T. 2024. *The Elephant and the Blind. The Experience of Pure Consciousness.* Cambridge, Mass.: MIT Press.

Metzinger, T. 2021. Artificial Suffering: An argument for a global moratorium on synthetic phenomenology. *Journal of Artificial Intelligence and Consciousness* 8(1), 43-66.

Metzinger, T. (ed.) 2000. *Neural correlates of consciousness*. Cambridge, Mass.: MIT Press.

Milner, D.A., Goodale, M.A. 1995. *The visual brain in action*. Oxford: Oxford University Press.

Mudrik, L., Boly,M., Dehaene, S., Fleming, S.M., Lamme, V., Seth, A., Melloni, L. 2025. Unpacking the complexities of consciousness: Theories and reflections. *Neuroscience & Biobehavioral Reviews* 170, 106053, doi.org/10.1016/j.neubiorev.2025.106053.

Nagel, T. 2013. *Mind and Cosmos*. Oxford: Oxford University Press.

Nagel, T. 2004. The psychophysical nexus. In: *Concealment and Exposure and other essays*. Oxford University Press, 194–235.

Nagel, T. 1974. What is it like to be a bat? *The Philosophical Review* 83, 435-450.

Nagel, T. 1971. Brain bisection and the unity of consciousness. *Synthese* 22 (3/4), 396-413.

Noë, A. 2009. *Out of our heads.* New York: Hill & Wang.

Noë, A., Thompson, E. 2004. Are there neural correlates of consciousness? *Journal of consciousness studies* 11, 3–28.

O'Regan, J.K., Noë, A. 2001. A sensorimotor account of vision and visual awareness. *Behavioral and Brain Sciences* 24(5), 883-917.

Papineau, D. 1998. Mind the gap. *Philosophical Perspectives* 12: *Language, Mind and Ontology*, pp. 373-89.

Parr, T. Pezzulo, G., Friston, K. 2022. *Active Inference. The Free Energy Principe in Mind, Brain, and Behavior.* Cambridge, Mass.: MIT Press.

Passos-Ferreira, C. 2024. Can we detect consciousness in infants? *Neuron* 112, 1520-1523.

Paul, L.A. 2014. *Transformative Experience.* Oxford: Oxford University Press.

Pereda, A.E. 2014. Electrical synapses and their functional interactions with chemical synapses. *Nature Reviews Neuroscience* 15(4), 250-263.

Pessoa, L. 2022. *The Entangled Brain. How Perception, Cognition, and Emotion are woven together.* Cambridge, Mass.: MIT Press.

Piccinini, G. 2021. *Neurocognitive Mechanisms. Explaining Biological Cognition.* Oxford: Oxford University Press.

Piccinini, G. 2009. Computationalism in the Philosophy of Mind. *Philosophy Compass* 4, 515-532. 10.1111/j.1747-9991.2009.00215.x.

Pitt, D. 2024. *The Quality of Thought*. Oxford: Oxford University Press.

Pitt, D. 2004. The phenomenology of cognition, or, what is it like to think that p? *Philosophy and phenomenological Research* 69(19), 1-36.

Putnam, H. 1965. The nature of mental states. In: *Mind, Language, and Reality. Philosophical papers Vol. 2*. Cambridge, Cambridge University Press: 429-440.

Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C. & Driver, J. (2000): Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain* 123(8:1), pp. 624–33.

Rensink, R.A., O'Regan, J.K., Clark, J.J. 1997. To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368–373.

Revonsuo, A. 2018. *Foundations of Consciousness*. London: Routledge.

Revonsuo, A. & Newman, J. 1999. Binding and consciousness. *Consciousness and Cognition* 8, 123–127.

Rosenthal, D.M. 2015. Quality Spaces and Sensory Modalities. In: *Phenomenal qualities: sense, perception, and consciousness*. Ed. by P. Coates, S. Coleman. Oxford: Oxford University Press.

Rosenthal, D.M. 2005. *Consciousness and Mind*. Oxford: Oxford University Press.

Sartre, J.P. 2003/1943. *Being and Nothingness. An essay on Phenomenological Ontology.* 2. Ed. London: Routledge.

Schlicht, T. 2025. Predictive Processing's Flirt with Transcendental Idealism. *Noûs.*

Schlicht, T. 2018. A methodological dilemma for investigating consciousness empirically. *Consciousness and Cognition* 66, 91-100.

Schlicht, T. 2017. Experiencing organisms. From Mineness to subject of experience. *Philosophical Studies* 175(10), 2447-2474.

Schlicht, T., Dolega, K. 2021. You can't always get what you want. Predictive Processing and Consciousness. *Philosophy and the Mind Sciences* 2: 8*.*

Searle, J.R. 1992. *The Rediscovery of the Mind*. Cambridge, Mass.: MIT Press.

Searle, J.R. 1980. Minds, Brains, and Programs. *Behavioral and Brain Sciences* 3(3), 417-457.

Seth, A.K. 2025. Conscious artificial intelligence and biological naturalism. *Behavioral and Brain Sciences.*

Seth, A.K. 2021. *Being You. A New Science of Consciousness.* London, Faber & Faber.

Seth, A.K., Bayne, T. 2022. Theories of consciousness. *Nat Rev Neurosci* 23, 439–452. https://doi.org/10.1038/s41583-022-00587-4.

Shea, N., Frith, C.D. 2019. The Global Workspace needs Metacognition. *Trends in Cognitive Sciences* 23(7), 560-571.

Simons, D.J., Chabris, C.F. 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception* 28(9), 1059-74.

Solms, M. 2021. *The hidden spring. A journey to the source of consciousness*. New York: Norton & Company.

Sperling, G. 1960. The information available in brief visual presentations. In: *Essential sources in the scientific study of consciousness*. Ed. by B. Baars et al. Cambridge, Mass: MIT Press 2003, 325-356.

Stanley, J., Williamson, T. 2001. Knowing (how). *The Journal of Philosophy* 98(8), 411-444.

Stubenberg, L., Wishon, D. 2023. Neutral Monism. *Stanford Encyclopedia of Philosophy* (Spring 2023 Ed.). Ed. by E. Zalta and U. Nodelman. https://plato.stanford.edu/archives/spr2023/entries/neutral-monism/.

Strawson, G. 2018. *Things that bother me. Death, Freedom, The Self etc.* New York: New York Review of Books.

Strawson, G. 2004. Real Intentionality. *Phenomenology and the Cognitive Sciences* 3(3), 287-313.

Strawson, G.1994. *Mental Reality.* Cambridge, Mass.: MIT Press.

Sur, M., Rubinstein, J.L.R. 2005. Patterning and Plasticity of the Cerebral Cortex. *Science* 310, 805-810.

Telakivi, P. 2023. *Extending the Extended Mind*. *From Cognition to Consciousness*. London: Palgrave.

Thompson, E. 2007. *Mind in Life.* Cambridge, Mass.: Harvard University Press.

Tononi, G. 2008. *Consciousness as integrated information: a provisional manifesto. Biol. Bull.* 215(3), 216–242.

Tsuchiya, N., Wilke, M., Frässle, S., Lamme, V.A.F. 2015. No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends in Cognitive Sciences* 19 (12), 757-770.

Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge, Mass.: MIT Press.

Van Gulick, R. 2004a. So many ways of saying no to Mary. In: P. Ludlow, Y. Nagasawa, and D. Stoljar (eds.), *There is something about Mary.* Cambridge, Mass.: MIT Press.

Van Gulick, R. 2004b. Higher-order global states. In: *Higher-order Theories of Consciousness*. Ed. by R. Gennaro. Amsterdam and Philadelphia: John Benjamins Publishing.

Varela, F.J., Thompson, E., Rosch, E. 1991. *The embodied mind*. Cambridge, MA: MIT Press.

Vold, K. 2015. The parity argument for extended consciousness. *Journal of Consciousness Studies* 22(3-4), 16-33.

Weierich, M.R., Wright, C.I., Negreira, A., Dickerson, B.C., Barrett, L.F. 2010. Novelty as a dimension in the affective brain. *NeuroImage* 49(3), 2871-2878.

Whitehead, A.N. 1920. *The concept of nature*. Cambridge: Cambridge University Press.

Zahavi, D. 2014. *Self and Other*. Oxford: Oxford University Press.