Abstract: 236 words
Main text: 7.106 words
Figures: 3

Consciousness as a Concrete Physical Phenomenon

Jussi Jylkkä*[1,2] & Henry Railo[2,3,4]

[1] Department of Psychology, Åbo Akademi University, Finland
Address: Fabriksgatan 2, 20500 Turku, Finland
E-mail: jjylkka@abo.fi

[2] Turku Brain and Mind Center, Finland

[3] Department of Clinical Neurophysiology, University of Turku and Turku University Hospital
Address: PL 52, 20521, Finland
E-mail: hmrail@utu.fi

[4] Department of Psychology, University of Turku, Finland

*Corresponding author

Abstract

Why any type of neural activation is associated with subjective, conscious experience is a fundamental unsolved question in neuroscience. To bridge the gap between neural activity and conscious experience, one seemingly must tie together two very different entities. The paradigmatic view in consciousness science is that subjective experiences are private and cannot be scientifically observed or described: science can only discover correlations between subjective experiences and their neural realizers, but never observe or describe the experiences themselves. We present a metatheory of consciousness that explains how subjective experiences are related to empirical observations and models, and why the two appear so different from each other. We argue that consciousness is a concrete physical process that causally interacts with other physical phenomena. This entails that consciousness can be empirically observed and characterized. The reason why subjective experiences and empirical observations and models of them appear so different is explained by what we call the observer-observed distinction. Empirical observations are always distinct from the observer, but a subject and her experiences constitute a single physical-biological system. We argue that once science has completely described 1) the constitutive neural mechanisms that are isomorphic with experiences, 2) the etiological mechanisms that experiences are based on, and 3) their causal power, then science has provided an exhaustive description of phenomenology. Our conclusion is that, if we accept this framework we call Naturalistic Monism, consciousness collapses into a standard problem of science.

Keywords: Constitutive mechanisms of consciousness, Empirical observation, Epistemic gap, Hard problem of consciousness, Neural correlates of consciousness, Phenomenology, Subjective experience

1. Introduction

The problem of consciousness is often considered as one of the biggest mysteries in science and philosophy. It is typically framed as follows: how can experiences, such as pain, emerge from physical phenomena in the brain. In fact, this "hard problem" of consciousness is so difficult that many philosophers and neuroscientists cannot agree as to how it could, even in principle, be answered. We argue that conscious experiences are concrete physical processes not categorically different from other physical phenomena; they have causal power and can be empirically observed and characterized. We propose that the "epistemic gap" between a subjective experience and its neuroscientific description can be reduced to the distinctness between the happening of the experience in a subject and how it is observed and characterized by an external observer. Our framework can be considered as a *metatheory* that clarifies in what sense empirical theories of consciousness explain subjective consciousness. We call this approach Naturalistic Monism.

The difficulty of conceiving subjective experiences—or "qualia", to use the philosophical term referring to the "qualitative feelings" associated with experiences—as an object of scientific study has led some empiricists to explicitly avoid referring to them (Dehaene, 2014), or to claim that such aspects do not exist (Dennet, 1991; Dennett, 2018). Others disregard the anti-physicalistic philosophical arguments and hold that the problem of consciousness is no different from other scientific mysteries: once the empirical science of consciousness has made sufficient progress, the emergence of experiences from neural activity loses its magical appearance, much in the same way as the replicating capacity of biological organisms stopped seeming magical after the discovery of DNA. This approach holds that the mystery of consciousness is not categorically different from other scientific problems. The problem with this kind of optimistic view is that it largely sidesteps the anti-physicalistic arguments instead of confronting them. Here, our attempt is to explain where the epistemic gap stems from, and show that experiences can be explained within a physicalist framework.

The empirical approach to solving the problem of consciousness is to determine the neural correlates of consciousness (NCC): the minimal set of neural processes that are together sufficient for a specific conscious experience (Crick and Koch, 1990). This approach is based on the assumption that an experience is identical with some neural processes. The grand objective of neuroscience can thus be said to be to uncover the "constitutive mechanisms of consciousness" (CMC; Revonsuo, 2006). Here, we use the term "CMC" instead of "NCC" to emphasize that empirical explanations of consciousness aim to characterize the causal physical processes that *are* consciousness (i.e. not processes that *cause* consciousness, or merely *correlate* with it). However, the problem posed by the anti-physicalist arguments by Chalmers and others is what it means for an experience and its constitutive mechanism to be identical, given that they appear so different. How can we suppose that a neuroscientific description of a red experience completely captures its nature, given that the feel of the experience cannot be inferred from its scientific description? Chalmers (1995) would argue that discovering the CMC is merely the "easy problem" of consciousness, whereas the "hard problem" is why there is an epistemic gap between an experience and its CMC.

The objective in the present article is to provide a metatheory for scientific studies of consciousness that specifies what it means for an experience E to be identical with its CMC_E, and why there is an epistemic gap between the two. Our framework of Naturalistic Monism is built on sharply distinguishing between concrete natural phenomena and our observations and theoretical characterizations of them. My consciousness, in all its qualitative richness, is a concrete natural phenomenon which I know simply because it *constitutes* my subjective realm—I am that phenomenon. Should the natural process that is my consciousness stop, I would no longer exist, for all that I'm concerned. My consciousness is the natural phenomenon that is me, whereas every other physical phenomenon is *distinct* from me. I can observe your consciousness, but this inevitably happens through causal interaction between distinct phenomena: your consciousness and my

consciousness. For you, your consciousness is just the *happening* of the process that I would describe, for example, as "a specific type of neural activation".

Naturalistic Monism implies that an empirical theory of the neural process CMC_E that directly corresponds to an experience E is an empirical theory *of the experience E*—the experience is the concrete phenomenon that external observers can scientifically describe as certain type of neural activation. The reason why the neural characterization of E seems so different from the subjective experience E itself is because, for the experiencer, E is part of the physical process that *constitutes* her consciousness, whereas for the external observer, E is a *distinct* physical process. This crucial difference between how experiences appear to the experiencer and how they appear to an external observer has, in our view, led many to falsely assume that experiences are somehow fundamentally different from other physical phenomena. There is no ontological difference between experiences and other physical phenomena, there is only a difference in point of view. We argue that, since experiences are concrete physical phenomena, empirical science can fully describe their nature.

The manuscript is divided into two main sections. In section 2, we present the general framework of Naturalistic Monism, and in section 3 we discuss the empirical implications of this view. But first, let us clarify some central theoretical terms.

1.1. Defining consciousness

Consciousness is extremely hard to define non-circularly, and instead of a detailed definition, researchers often give practical examples of what they mean by the term. Descriptions of consciousness vary from person to person, and all definitions are distinct from the concrete phenomenon they aim to describe. If you take René who believes in an immaterial soul, David who believes in qualias, and Daniel who is an eliminativist, and stick them with a needle, the same kind of natural process is instantiated in all of them, and it feels very similar to each of them. To avoid

the trap of falsely defining consciousness, I could define "consciousness" *ostensively* as this; the totality of my subjective reality right now (cf. Ismael, 1999).

According to a famous conceptualization provided by Block (1995), phenomenal consciousness should be separated from access consciousness which refers to the functional aspects of consciousness. We do not make such distinction. On our view, consciousness is a causal process in the brain that is in constant interaction with other physical processes, some of which may lead to overt behavior, others not. In this sense, consciousness is always "functional". We agree with Dehaene that postulating the existence of purely phenomenal consciousness that is distinct from conscious access is "highly misleading and leads down a slippery slope to dualism" (p. 10, Dehaene, 2014). Yet, associating consciousness completely with "access" is likewise problematic because it leads one to disregard the experiential nature of consciousness (Carter et al., 2018). On our approach, as explained below in more detail, consciousness is a physical process that, when it occurs in a conscious subject, can be described by the subject as "feeling like this", but which is always also causally efficacious.

2. Naturalistic Monism

Naturalistic Monism consists of an ontological and an epistemological component. The ontological claim is simply that consciousness is a concrete physical process in the world in the same sense as lightning, photosynthesis, and life are, and can be empirically observed and characterized. We take this to be commonsense physicalism. The problem for this view is the epistemic gap. Next, we will argue that the epistemic gap due to what we call the observer-observed distinction. This is the claim that consciousness is a concrete phenomenon in the world, whereas observations and theories are distinct from it.

The reason why our consciousness seems so different from phenomena like neural activity is because consciousness is itself a concrete physical phenomenon, whereas terms like "neural activation" are theoretical constructs that we use to *model* concrete phenomena in the world, based on empirical observations *caused* by the concrete phenomena in us.[1] All empirical science is based on systematic, theory-driven observations. This is fundamentally different from the way we know our consciousness. Consciousness is not distinct from me, it is not something I observe through causal interaction. Instead, when I am conscious of, say, pain, the pain is not distinct from me; rather, it is part of the physical system that constitutes my consciousness. This is why I cannot doubt the existence of my pain like I can doubt the existence of "external" phenomena (i.e., phenomena that do not constitute my consciousness).

The fact that the observer-observed distinction does not hold in the case of consciousness does not imply that consciousness would not be a physical phenomenon, nor even that it could not be observed empirically by an external observer. *All* concrete physical phenomena can be observed, because they can causally interact with an observer. The physical phenomenon in the world which is an experience E of a subject S is experienced from the "first-person perspective" (as consciousness, or *this*) by S, because it is part of the physical process that constitutes the consciousness of S. This physical process E can also be observed empirically by an external observer, who can model E as, say, complex neural activity of type X. A similar case is made by Edelman and Tononi (2000), who argue that "Unlike any other entity, […] with consciousness *we are what we describe* scientifically" (p. 14, italics in the original). The reason why our subjective consciousness seems so different from other physical phenomena is because physical phenomena that are *not* part of our consciousness are always only observed empirically (that is, through causal interaction between an observer and the

[1] A similar point is made in physics by Hawking and Mlodinow (2010), who put forward a view they call model-dependent realism. This view implies that we can never directly know the world as it is in itself, but instead we can build models of the world based on sensory input. We hold that the relationship between an experience and its constitutive mechanisms is similar: the constitutive mechanism is a theoretical model of an experience, which is a concrete phenomenon in the world.

observed). A phenomenon that I empirically observe is always distinct from the physical phenomenon that is my consciousness.

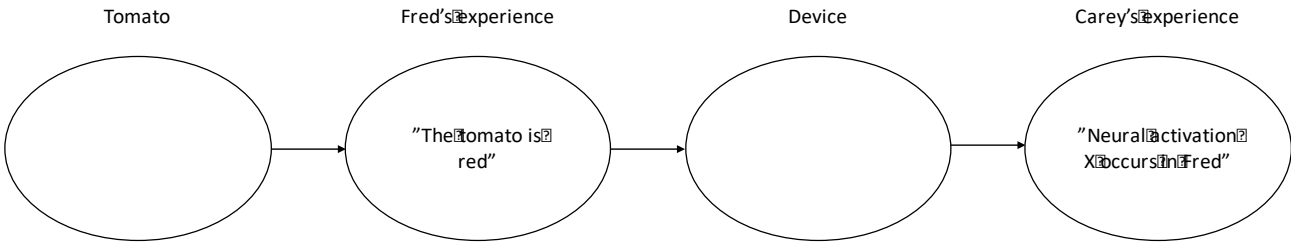

Figure 1. *Causal interaction between four physical phenomena*

To illustrate, consider the interaction between four physical phenomena: a red tomato, an experience of the tomato's color in Fred, a scanning device that can detect the constitutive mechanism of Fred's color experience, and a neuroscientist Carey examining Fred's color-experience with the device. This process is illustrated in Figure 1, where all the ellipses represent physical phenomena. The texts in quotation marks inside the ellipses represent how Fred and Carey might verbalize their experiences. The tomato causes an experience of red in Fred, which is detected by the scanning device. The readings of the scanner are observed and interpreted by Carey as, say, neural activation X. Carey's interpretation of the scanner readings as "neural activation X" is a theoretical model of what causes the scanner readings. For Fred, the happening of the phenomenon that Carey represents as "neural activation X" is his experience of red. The reason why Fred's experience and Carey's representation of it as "neural activation X" appear different is because they *are* different: Fred's and Carey's experiences are two distinct physical phenomena, caused by two distinct phenomena.

This does not imply that Carey's characterization of Fred's experience as "neural activation X" could not be an exhaustive *description* of the experience. Our representations of external phenomena are always very different from what they represent. For instance, a person with intricate knowledge of atomic physics could perfectly describe and represent in her mind the splitting of an

atom, but the actual happening of that phenomenon is very different from the representation of it—the concrete phenomenon releases a huge amount of energy, whereas the representation does not. Similarly, Fred's experience feels like red, whereas Carey's representation of it does not.

The epistemic gap is often formulated in terms of what experiences feel like. So, what *do* experiences feel like? What an experience feels like is impossible to capture in any descriptions—for instance, a person with congenital achromatopsia cannot know what red feels like no matter how we describe it to her. For anyone to know what an experience feels like requires the *happening* of the experience in her, the occurrence of the concrete phenomenon. I can say that my experience of the music I am currently listening feels like *this*, where the term "this" refers to my current experience of the music. In general, we can give the following interpretation for any statement about what an experience feels like:

(Q1) What an experience E feels like for a subject S is the happening of E in S.

The experience E is a physical process that can be empirically described as, say, neural activation X. The terms "neural activation X" and "experience" are coextensive, and we can substitute one for the other. Thus, the statement (Q1) can be reformulated as follows:

(Q2) What neural activation X feels like for a subject S is the happening of neural activation X in S.

There is nothing in these statements that conflicts with physicalism: physicalism consists of the claim that all concrete phenomena are physical, and can be scientifically *described*. A description of a concrete phenomenon, no matter how accurate and detailed, can never by itself cause its happening. This does not imply that the description could not fully and exhaustively capture the

nature of that phenomenon. For instance, if an experience E consisted solely of causal interaction of type T between neural modules M1 and M2, then we could fully describe the experience with the description "T-interaction between M1 and M2". But this is just a *description* of the phenomenon whose happening in a subject S constitutes its what-it-is-likeness for S. The take-home message is that we need not postulate any terms for the "qualitative properties" of the experience E in its description over and above the terms "type T interaction", "M1", and "M2": when the process *described by these terms* happens in S, it constitutes its "feel" for S. At this point, one might object that the scientific description cannot fully capture the nature of the experience, because for S the experience does not introspectively *appear* as "type T interaction between M1 and M2". However, such an argument is flawed: from the fact that an empirical description exhaustively describes an experience E, it does not follow that the subject S experiencing E would introspectively see that such a description captures the nature of her experience (Dennett, 2018).

This reconceptualization of the epistemic gap can be illustrated with the Mary thought experiment. When Mary is in the black-and-white room, she knows everything about color experiences empirically; she fully and exhaustively grasps the scientific *descriptions* of color experiences. When she leaves the room, she does not gain new *descriptive* knowledge, but instead *a completely new physical process is instantiated in her brain*—a concrete color experience whose scientific descriptions she already knew perfectly well. Mary learns "what it is like" to see colors, and in line with the Q-claims above, this should be reinterpreted as the claim that color experiences *start to happen* in her. This does not contradict physicalism, which only claims that all physical phenomena can be *described* scientifically.

2.1. Relationship between an experience and its constitutive mechanism

Naturalistic Monism implies that an experience is a concrete physical process that can be empirically modelled. Thus, the terms "experience of red" and "the constitutive mechanism of red

experiences" ("CMC_{red}") refer to the same concrete process (type[2]) in the world. Whereas the term "experience of red" is typically used to refer to an experience when it is part of a subject's consciousness, the neuroscientific term "CMC_{red}" is used in scientific contexts, where it refers to the experience as the external phenomenon that causes scientific observations. Experiences are complex physical phenomena that are based on lower-level phenomena, and the basedness can be described scientifically. This framework is summarized in Figure 2.

We clarify our view with a neuroscientific example. Assume that a given process of the brain CMC_{cat} always corresponds to a subjective visual experience of a cat (be it in a waking state, dream, or hallucination). This process could consist, for instance, of sustained recurrent activation between subcortical areas (such as the thalamus) and specific type of interaction of frontal, parietal and content-specific sensory cortical areas. Whenever anyone's brain realizes this process, they experience seeing a cat. The CMC_{cat}-theory is built through empirical observation, by determining which neural mechanisms causally lead to a subjective experience of a cat. The observations that we have used to build the theory—measurements of neural activity, reports of conscious perception of a cat, and so on—are distinct from the phenomenon itself, but caused by it. According to Naturalistic Monism, the concrete physical process in the world that the CMC_{cat}-theory empirically characterizes *is* an experience of a cat. The CMC_{cat}-theory can exhaustively *describe* the nature of this process, although it does not, and cannot, afford knowledge of its what-it-is-likeness, which is the *happening* of the process in a subject (cf. the Q-claims above). When the physical process that the CMC_{cat}-theory describes occurs in *my* brain, that process is my subjective experience. Put differently, "experiences", or "consciousness" more generally, refers to certain observer-

[2] Philosophers like to distinguish between types of phenomena (the type of experiences that are red) and its instantations, or "tokens" (a specific red experience). Both can be characterized empirically, but because in science one always aims to make generalizations, it is arguably more important to specify the mechanism for certain types of experiences rather than their specific instantiations.

independent physical processes, and the aim of empirical science is to characterize these physical processes using observations.

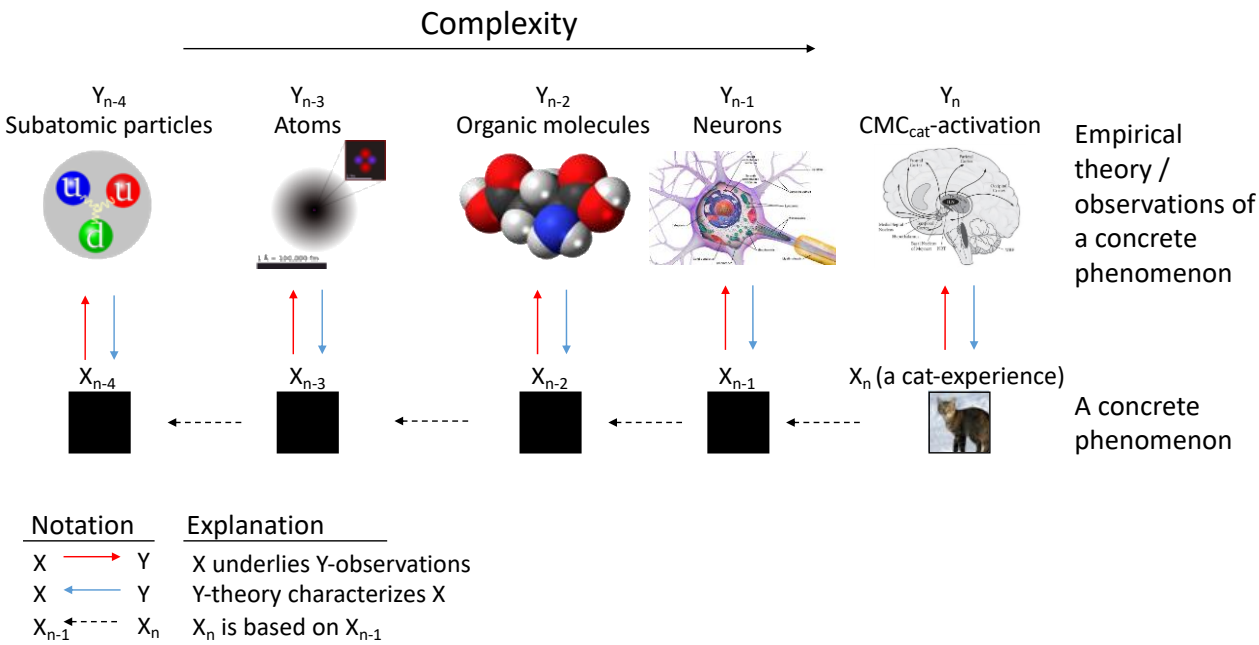

Figure 2. *A schematic presentation of Naturalistic Monism. Empirical observations (upper row) are caused by concrete physical phenomena (red arrows), and empirical theories characterize the concrete physical phenomena (blue arrows). Complex physical phenomena are constituted by more basic level phenomena (dashed arrows). Experiences, as depicted by the picture of a cat, are concrete physical phenomena, and hence can be empirically observed and described by scientific theories. The lower-level constituents of experiences are shown in black because they are not part of consciousness; they can only be characterized empirically. (All images under Creative Commons license.)*

The cat-experience is depicted in Figure 2 as a complex physical process that is based on lower-level constituent processes. This basedness can be empirically clarified by defining how the CMC_{cat}-process is related to its constituent processes. It is sometimes argued that consciousness cannot be fully explained by analyzing the interactions between its lower-level constituents. This argument conflates concrete phenomena with their descriptions: one can never "reduce" any concrete phenomenon to lower level processes, because reduction is always explanatory. One can *explain* how consciousness, for instance an experience of a cat (CMC_{cat}), is based on lower level processes, but this does not amount to "reducing" the concrete experience into anything else. We can explain (in terms of CMC-descriptions) how an experience is based on lower-level processes,

but that does not amount to reducing *the concrete process* into anything else. If we would tear the process itself into parts, it would cease to exist.

The black boxes in Figure 2 might appear to motivate some kind of mysticism about the nature of non-mental physical phenomena, similar to Russellian "intrinsic nature" that cannot be empirically captured. We agree that there is a *pinch* of mystery here, but nothing that would contradict science or physicalism. Although we cannot know anything about the nature of an external phenomenon without relying on observation or theory (that is, without representing it in some way in our consciousness[3]), we know that the external phenomenon is such that it causes our observations, and that our theories model it. We do not postulate the existence of properties that are empirically unobservable; we simply hold (trivially) that any concrete phenomenon exists independently of our observations and theories of it. The reason why all the other boxes in Figure 2 besides the cat experience are black is because none of the other phenomena are *identical* with the experience (none of the other processes are CMC_{cat}). Our experiences are the only concrete phenomena in the universe that constitute our phenomenal realm. We do not know their nature only empirically, we *are* that phenomenon.

3. Towards an empirical theory of phenomenology

Naturalistic Monism has following implications for empirical science:

1. There is no "hard problem" to be explained. Consciousness is a concrete physical phenomenon that can be measured, and explained by science.

[3] The classical question "What is the sound of a tree falling when no one hears it?" can be translate into more scientific terminology by asking "What is the nature of an electron if you do not utilize observations or theory to represent it in your mind?" We know that electrons exist—otherwise there would be nothing to cause our electron observations—but we do not know what an electron is without representing it in our mind in some way or another. Hence the black boxes in Figure 2.

2. Isomorphism. A scientific theory CMC_E of experience E is, literally, an empirical description of the phenomenology of experience E.
3. Constituents. Science aims to describe what kinds of lower-level processes constitute consciousness.
4. Etiology. Understanding the phylogenetic and otogenetic basis of consciousness is crucial for understanding phenomenology.
5. Causal power. Because consciousness is a concrete physical phenomenon, it can causally directly interact with other physical processes.

Based on these, we argue that empirical science can fully explain the phenomenology of experience. This is summarized in Figure 3. We will next examine each of the above claims in a bit more detail.

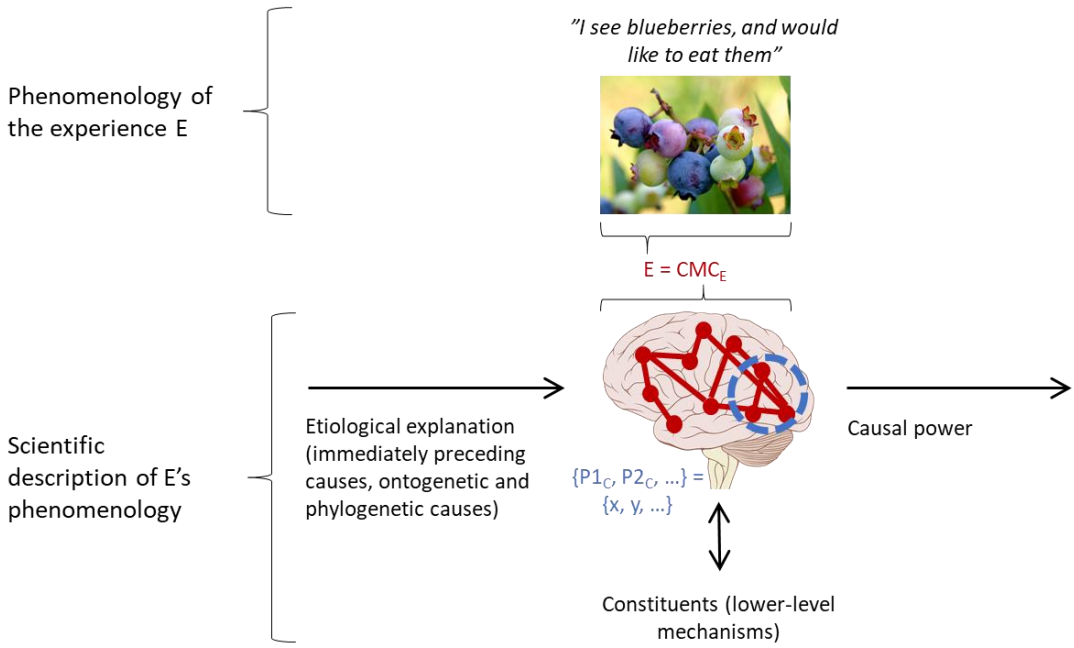

Figure 3. *Understanding phenomenology empirically. Phenomenology of experience E is depicted in the figure by the picture of blueberries. We assume that at this moment the subject experiences seeing blueberries, feels an urge to eat them, and says "I see blueberries, and would like to eat them". The scientific model of E is here depicted by the image of the brain (CMC_E), where color is represented as a set of color parameters {$P1_C$, $P2_C$, ...} that can have different values {x, y, ...}. Our key claim is that the scientific CMC_E-description is a description of the experience E. Hence, CMC_E and E, although appearing to be different (due to the observer-observed distinction), are isomorphic, and E can be described by a detailed empirical model (CMC_E). In addition to explaining this isomorphic relation between experiences and empirical models of them, we propose that to understand phenomenology, empirical science must describe how the CMC_E is constituted by lower level processes, and give an etiological explanation that consists of a description of the immediately preceding causes of E, as well as its ontogenetic (developmental), and phylogenetic*

(evolutionary) antecedents. Causal power here refers to the neural processes and behavior that the experience E (empirically described as CMC_E) causes. [All images under Creative Commons license.]

3.1. There is no "hard problem" to be solved

If we accept that experiences are concrete physical phenomena in the brain, then the task of a neuroscientific theory of consciousness is to characterize those phenomena neuroscientifically. Put differently, the aim of the empirical scientific theories of consciousness becomes to solve what has been called the "easy problems" of consciousness (Chalmers, 1995), because the problems can be solved using the standard scientific approach.

Our approach helps empirical theories of consciousness to avoid postulating arbitrary psychophysical laws that specify how phenomenal properties of the world depend on physical properties (Chalmers, 1996). According to Naturalistic Monism, an empirical theory of consciousness is exactly what its name says: it is a theory of conscious physical phenomenona that is based on empirical observations of those phenomena. The concrete phenomenon in the world that the theory models as, say, information processing, higher-order processing, global workspace activation, recurrent activation, and so on, *is* consciousness—when it happens in me it is *my* consciousness, and feels like *this*. There is no ontological gap to be bridged between mental and non-mental phenomena.

3.2. Isomorphism

An empirical theory of consciousness, if sufficiently detailed and accurate, can provide a complete scientific description of an experience. Isomorphism refers to the fact that there is a perfect one-to-one correspondence between the empirically observed phenomenon (CMC_E) and the subjective experience E (Fingelkurts & Fingelkurts, 2004; Revonsuo, 2000).

According to the present framework, a neuroscientific explanation of color experience should elucidate the isomorphic relationship between colour experience and neural activity; the aim is to model how interactions in the networks of neurons correspond to experienced color.[4] If a neural process X is isomorphic with experienced color, then an empirical model that describes X effectively describes color phenomenology—the concrete process X constitutes the color experience. This type of model of color phenomenology is depicted in Figure 3 as the set of color parameters. The example is obviously hugely simplified, and it is likely that color phenomenology cannot be completely accounted for by simply activity in single visual cortical area. Moreover, color is typically not perceived alone, but it is seen as part of an object which is located in three-dimensional space (Witzel & Gegenfurtner, 2018), and color phenomenology could be shaped by language (Maier & Rahman, 2018).

According to our framework, when a neuroscientist examines an experience of blue in a subject S, the phenomenon that causes the neuroscientist's CMC_{blue}-observations and that her CMC_{blue}-model describes *is* the experience of S. Any change in S's experience of blue corresponds to a change in the neuroscientist's CMC_{blue}-observation (because it is caused by the the experience), and the neuroscientist can model changes in S's experience by the parameters in the CMC_{blue}-model. In short, because the CMC_{blue}-model is a description of S's red experience, it is also a description of its phenomenology. More complicated experiences could be described in a similar fashion using more complicated models (e.g. Oizumi et al., 2014).

[4] Consciously perceived colors form a color space where adjacent colors are separated by a similar amount ("just noticeable differences"). This can be measured using psychophysics (Munsell, 1919). The color space human experience does not simply reflect the wavelength of the light. The objective spectrum of light is linear, whereas the subjective color space we experience is circular, and divided to color categories that clearly differ from each other (red, green, blue, yellow).

The reader may, at this point, insist that such physicalistic description would not explain the "qualitative feeling" of experiences, or answer why blue feels like blue, and hence the CMC_{blue}-description would not be an explanation of what blue feels like. Such counter-arguments are ill-posed, because they conflate concrete phenomena with their descriptions. What blue feels like for S is the happening of CMC_{blue} in S, a concrete process that can be described scientifically. Of course, grasping the CMC_{blue}-description does not afford "knowledge" of what blue feels like, because what blue feels like for S is the happening of CMC_{blue} in S (cf. the Q-claims in section 2). No description can ever cause the happening of what it describes, but this does not imply that it could not perfectly describe the phenomenon and its happening in a subject[5].

3.3. Constituents

According to Naturalistic Monism, a successful empirical theory of consciousness must specify the *constituent processes* that are necessary for consciousness, but which are not themselves conscious until they interact in some specific way to form the CMC. In a similar vein, hydrogen and oxygen are constituents of water, but they are not water until they bind in a specific manner. The aim of empirical science of consciousness is to describe what needs to happen so that the constituent processes "cross the threshold" to consciousness, in the same sense that physics and chemistry characterize how hydrogen and oxygen are combined to form water (instead of hydrogen peroxide,

[5] As to the question of why blue feels like blue, it is important to keep in mind that the subjective experience of blue is a concrete natural phenomenon. Asking why blue feels like blue is like asking why helium has two protons in its nucleus. If helium did not have two protons, it would not be helium, and if blue did not feel like blue, it would be another natural phenomenon. However, just like science can describe how helium was formed shortly after the Big Bang through the interaction of free subatomic particles, science can explain how blue experiences have been formed in evolution as an adaptation to efficiently discriminate external stimuli. Philosophical questions such as why CMC_{blue} is associated with a subjective experience of blue instead of, say, red, are ill-formed in the light of the presented framework. The scientific CMC_{blue}-theory is about the natural phenomenon blue experience, not any other phenomenon, just like the helium-theory is about the natural phenomenon helium and not hydrogen.

H_2O_2).[6] This is exactly what the current search for the NCC is about (e.g. Dehaene & Changeux, 2011; Crick & Koch, 1990; Koch et al., 2016).

The constituents determine the phenomenology of consciousness – they are what experiences are "made of". The constituents include anatomical (e.g. brain areas, and their connections) and physiological (e.g. neural activity) factors. The primary visual cortex (V1) can be said to constitute normal conscious visual perception: it is an essential node through which subcortical and cortical visual areas communicate, and hence lesions to it produce blindness (Tong, 2003). A more sophisticated example about how V1 may contribute to visual phenomenology is that the surface area of V1 predicts individual differences in the experienced size of visual stimuli (Schwarzkopf, Song, & Rees, 2011). Visual content in general – for example, whether you are seeing faces or houses – is determined by which areas are active (Tong, Nakayama, Vaughan, & Kanwisher, 1998). In conclusion, at any moment, the phenomenology of conscious vision is based on the activity of a wide network of areas; the experience is literally made of its constituents.

In addition to (exteroceptive) sensory contents, consciousness can include interoceptive contents; emotions; feelings of agency, motor as well as cognitive control; memories; language; self-consciousness, and so on. According to Naturalistic Monism, each of these contents of consciousness are concrete physical processes that can be empirically observed and modelled. It is an open empirical question whether conscious experiences always build on some "core processes" such as metacognition (Lau & Rosenthal, 2011), memory (Edelman, 1989), representation of space

[6] Note that the constituent processes of consciousness do not *cause* consciousness, just as hydrogen and oxygen do not cause water. Instead, when the constituent processes and bound in a specific manner, they form a new process, which has properties and causal powers that the constituents lack. However, such properties are not "emergent" in the strong sense of the word, because they are directly based on how the constituents are bound. Neuroscience can explain constitution in the case of consciousness just like physics and chemistry can explain how water is constituted from hydrogen and oxygen.

(Merker, 2007; Revonsuo, 2006), and/or interoceptive signals (Park & Tallon-Baudry, 2014; Seth, Suzuki, & Critchley, 2012).

As a practical example of the constituent mechanisms of consciousness, consider the global neuronal workspace theory (Dehaene & Changeux, 2011). The proponents of the theory would argue that consciousness is constituted by synchronous interactions between specific neural populations in frontal, parietal and sensory cortical areas. This CMC is made of its constitutive processes such as specific cortical areas, neuron types, and activity patterns (e.g. synchronous oscillations at specific frequencies); these determine the content of experience. According to the theory, these constituents are not conscious alone, but become conscious when they interact in the way that the theory describes. These constituent processes are, of course, further based on lower-level processes: oscillatory activity requires a certain balance of excitatory and inhibitory neurons, neural activity requires a resting membrane potential, neurons are made up of specific molecules, and so on. The constituents can, in principle, be traced to fundamental physical entities; in this sense, the CMC (the experience) is continuous with other physical phenomena.

3.4. Etiology

In addition to characterizing what types of processes constitute consciousness (CMC), empirical theories of consciousness can describe the etiology that has led to consciousness and its specific contents (Revonsuo, 2006). Naturalistic Monism implies that etiology can help to explain not only the functions of consciousness, but also its phenomenology. Here, following Revonsuo (2006), we separate three classes of etiological explanations of consciousness: 1) immediately preceding causes, 2) ontogenetic explanations, and 3) evolutionary explanations.

Immediately preceding causes refer to processes that happen right before a content enters consciousness. Conscious perception of color of external objects causally dependes on the spectrum

of the light that hits the retina, the photoreceptor cells in the retina (which are sensitive to different wavelengths of light), color-opponent cells along the visual pathway (which code opposite-signed colors, e.g. red vs. green), and cortical areas such as the V4 (Gegenfurtner & Kiper, 2003; Witzel & Gegenfurtner, 2018). These processes, in part, explain phenomenology, although they may not necessarily constitute it (e.g. retinal photoreceptors contribute similarly to conscious and unconscious perception). In a similar fashion, the quest to describe the correlates of consciousness attempts to describe the time-course of neural processes that lead to conscious perception (Dehaene & Changeux, 2011; Railo, Koivisto, & Revonsuo, 2011). According to the present account, the processes that lead to conscious perception and affect its contents are part of its etiology, and these processes, in part, determine phenomenology. For instance, the state of cortical neurons influence whether a subsequently presented visual stimulus will be perceived or not (Busch, Dubois, & VanRullen, 2009; Samaha, Gosseries, & Postle, 2017).

Second, in addition to immediately preceding causes, previously encountered information and individual development, in general, shapes the brain and consequently our experiences. Perceived color, for example, is influenced by prior experiences: you know that bananas are likely to be yellow, which biases your perception of their color (Lafer-Sousa, Hermann, & Conway, 2015; Witzel, Valkova, Hansen, & Gegenfurtner, 2011). In addition to merely biasing what kind of "inferences" the brain makes about the external world, ontogenetic development literally constrains the constituents of consciousness. Because individual development directly shapes the physiology and anatomy of the brain, it on the one hand limits, and on the other hand expands, potential experiential contents. The brain learns to experience contents that it encounters; an infant learns to perceive the outside world and her own body through trial and error. As an example, hearing a certain language during childhood will "tune" the auditory system to the phonemes present in that language (Näätänen et al., 1997). This is why we have trouble hearing the phonetic details of foreign languages.

Third, not only is the brain and the constituents of consciousness shaped by the individual's own history, but it is dictated by our genetics and phylogenetic development. This means that the etiology of phenomenology can be traced down to evolution. A complete biological theory of consciousness should explain how consciousness has evolved from earlier evolutionary products. For instance, pain can be considered as a complex biological adaptation that has helped organisms to avoid potentially damaging stimuli, and whose bio-chemical basis could possibly be traced back to single celled organisms (Bray, 2009). According to Naturalistic Monism, this does not mean that only the functional aspects of pain would be products of evolution, but also that the different types of subjective unpleasant sensations we call “pain” are products of evolution. A similar approach can be applied to other contents of consciousness. Consider vision: why do we experience visual stimuli the way we do? What we consciously see can be explained in terms of behavioral utility; neural circuits that allow useful discriminations, and thereby useful behavioral responses will be passed on in the evolution. Evidence for this comes from studies showing that the frequency a stimulus has been encountered in the environment predicts phenomenology (see, Purves et al., 2011). For instance, consciously perceived brightness of a visual target on some background, can be predicted by specifying how often that combination occurs in the real-world (Yang & Purves, 2004; the same has been observed for orientation and motion, see Purves et al., 2011). This shows that the organism’s ontogenetic and phylogenetic history help to explain why experiences feel like they do. In other words, these type of scientific, etiological descriptions explain phenomenology. This view also puts emphasis on the functional aspects of experiences; phenomenology is closely associated with behavioral utility.

3.5. Causal power

Naturalistic Monism ties together the phenomenal and functional aspects of consciousness. Because consciousness is a physical process, it can causally interact with other phenomena in the brain and

drive the organism's behavior. We propose that Naturalistic Monism may help to dismantle some stubborn conflicts about whether consciousness should primarily be described in functional or phenomenal terms. While it is an empirical question to find out whether specific cognitive functions (e.g. global access or error detection; Dehaene, Lau, & Kouider, 2017) are necessary for consciousness, our framework implies that experiences never take place in a causal void. An experience E (CMC_E) always causally interacts with other processes in the brain, both conscious and non-conscious: it can cause, for instance, thoughts, emotions, bodily states, or actions, and interact with the outside world. Behavioral processes such as reports are simply one of the many processes E can cause. Any experience E can be portrayed both through phenomenal and functional terms. It is important to note that both of these are still merely *descriptions;* neither of them is the concrete physical phenomenon which they refer to.

To repeat, the present framework considers experiences as physical phenomena that can influence other brain processes, cognition and behavior. Thus, contents of consciousness such as color, pitch, pain, and so on can be seen essentially as *discriminations* that the brain makes about the outside world; they are sources of information for the organism's behavior. Color, for instance, provides information about objects, their parts, and textures which enables faster and more accurate actions in the world, and more efficient communication with other people (Witzel & Gegenfurtner, 2018). Such "functionalistic descriptions" can give insight into phenomenology as well. For instance, according to the theory proposed by Purves and Lotto (2002), the "shape" of the color space humans experience arises from the need to discriminate colors in optimal fashion[7]. Similar logic can

[7] First, the experienced color space must appropriately reflect similarities and differences between physical spectra in a systematic fashion. Second, that humans experience four major classes of color (red, green, blue and yellow) may be the consequence of the so-called four-color problem: four colors are needed to disambiguate any complex set of overlapping objects which differ in their spectral properties (Purves, Lotto, & Polger, 2000). The conclusion, if the theory is correct, is that color phenomenology is a solution to the problem of discriminating objects (independently of luminance), which in turn allows more efficient behavior.

be applied to other contents of consciousness (Purves, Wojtach, & Lotto, 2011). This type of view speaks against the philosophical notion that experienced color (color "qualia") is epiphenomenal (Jackson, 1982), and supports the idea that color is the way the brain represents information, grounded in the way the brain extracts information from incoming stimuli.

Because according to Naturalistic Monism consciousness is a causal physical process, the aim of empirical science of consciousness is to describe how the causal physical process that is consciousness (CMC) can be differentiated from other processes that take place in the brain. To pinpoint the CMCs of specific conscious contents such as blueness, empirical studies of consciousness attempt to disentangle the processes that directly correspond to the experience of blue from the actions elicited by it (e.g. blueberry-picking, or in the case of empirical studies, reports of what a participant experienced; Tsuchiya, Wilke, Frässle, & Lamme, 2015). The reader may, at this point, argue that because science has shown that a behavior can also be guided by unconscious information[8], any "functionalistic description" will not be sufficient to describe consciousness. Here, it is important to note that the causal power of unconscious information on behavior is very limited compared to conscious perception (see, Revonsuo, Johanson, Wedlund, & Chaplin, 2000, for discussion about the empirical possibility, and potential empirical examples of zombie behavior). One of the goals of empirical science of consciousness is to explain what gives

[8] For instance, some of blindsight patients, who are blind in certain parts of the visual field due to a lesion in primary visual cortex, can nevertheless use chromatic information to influence their behavior (Stoerig & Cowey, 1992). Similarly, unconscious perception of chromatic information has been demonstrated in neurologically healthy observers, for instance, using unconscious priming methodology (e.g. Breitmeyer, Ro, & Singhal, 2004). However, recent research has also shown that previous studies have overestimated the power of unconscious processes to influence behavior. In early studies (in blindsight patients and neurologically healthy observers), stimuli that elicited some weak conscious percepts were incorrectly classified as "unconscious" because conscious perception was operationally defined crudely (Newell & Shanks, 2014; Overgaard, 2011; Peters & Lau, 2015). Moreover, in blindsight, unconscious perception is often associated with some conscious feelings about the stimulus (Ffytche & Zeki, 2011).

conscious processes their power to potentially elicit fast and accurate actions, and flexible voluntary behavior. Theories such as the Global Workspace theory are explicitly built around this foundation (Dehaene & Changeux, 2011). According to Naturalistic Monism, such "functional explanations" can provide key insights into the phenomenal qualities of experiences as well, because they are literally models of experiences.

4. Conclusions

We have argued that consciousness and experiences are concrete physical phenomena that causally interact with other physical phenomena, including observers. They can be empirically studied, and empirical science can discover their nature and show how they are based on lower level physical phenomena. We have argued that the epistemic gap can be reduced to the distinctness between concrete experiences and their theoretical characterizations, and that it does not pose a problem for physicalism and the scientific study of consciousness. Accordingly, neuroscientific theories of consciousness need not answer the "hard problem" of consciousness, which is based on conflating concrete phenomena with scientific models of them. We have argued that our approach helps empirical theories of consciousness to avoid postulating ad hoc psychophysical laws, as well as to reconcile the functional and phenomenal aspects of consciousness. The proposed framework helps us to get rid of dualistic thinking, which has arguably hindered considering consciousness as a true physical phenomenon.

5. Acknowledgements

J.J. was supported by the Abo Akademi University Endowment (the BrainTrain project and personal grant) and H.R. by the Academy of Finland (grant #308533). We thank Antti Revonsuo, Valdas Noreika, and Adam Bricker for helpful comments and discussions.

6. Data availability

No data was gathered for this study.

7. References

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X00038188

Bray, D. (2009). *Wetware: A Computer in Every Living Cell*. Yale: Yale UP.

Breitmeyer, B. G., Ro, T., & Singhal, N. S. (2004). Unconscious Color Priming Occurs at Stimulus-Not Percept-Dependent Levels of Processing. *Psychological Science*, *15*(3), 198–202. https://doi.org/10.1111/j.0956-7976.2004.01503009.x

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The Phase of Ongoing EEG Oscillations Predicts Visual Perception. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0113-09.2009

Carter, O., Hohwy, J., Van Boxtel, J., Lamme, V., Block, N., Koch, C., & Tsuchiya, N. (2018). Conscious machines: Defining questions. *Science*. https://doi.org/10.1126/science.aar4163

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

Dehaene, S. (2014). *Consciousness and the Brain - Deciphering How the Brain Codes Our Thought*. New York: Penguin Books.

Dehaene, S., & Changeux, J. P. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*. https://doi.org/10.1016/j.neuron.2011.03.018

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science*. https://doi.org/10.1126/science.aan8871

Dennet, D. . (1991). *Consciousness Explained*. Boston: Little, Brown and Co.

Dennett, D. C. (2018). Facing up to the hard question of consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), 20170342. https://doi.org/10.1098/rstb.2017.0342

Edelman, G. (1989). *The Remembered Present: A Biological Theory of Consciousness*. Basic Books.

Edelman, G. M., & Tononi, G. (2000). *A Universe of Consciousness: How Matter Becomes Imagination*. Basic Books.

Ffytche, D. H., & Zeki, S. (2011). The primary visual cortex, and feedback to it, are not necessary for conscious vision. *Brain*. https://doi.org/10.1093/brain/awq305

Fingelkurts, A. A., & Fingelkurts, A. A. (2004). Making complexity simpler: Multivariability and metastability in the brain. *International Journal of Neuroscience*. https://doi.org/10.1080/00207450490450046

Francis Crick and Chri.dof Koch. (1990). Towards a neurobiological theory of consciousness. *The Neurosciences*. https://doi.org/10.1109/INEC.2010.5424508

Gegenfurtner, K. R., & Kiper, D. C. (2003). C <scp>OLOR</scp> V <scp>ISION</scp>. *Annual Review of Neuroscience*. https://doi.org/10.1146/annurev.neuro.26.041002.131116

Hawking, S., & Mlodinow, L. (2010). *The Grand Design*. New York: Bantam Books.

Ismael, J. (1999). Science and the phenomenal. *Philosophy of Science*, *66*(3), 351–369.

Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly*, *32*(127), 127–136.

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn.2016.22

Lafer-Sousa, R., Hermann, K. L., & Conway, B. R. (2015). Striking individual differences in color perception uncovered by “the dress” photograph. *Current Biology*. https://doi.org/10.1016/j.cub.2015.04.053

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2011.05.009

Lotto, R. B., & Purves, D. (2002). A rationale for the structure of color space. *Trends in Neurosciences*. https://doi.org/10.1016/S0166-2236(02)02059-3

Maier, M., & Rahman, R. A. (2018). Native Language Promotes Access to Visual Consciousness. *Psychological Science*, *29*(11), 1757–1772. https://doi.org/10.1177/0956797618782181

Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X07000891

Munsell, A. H. (1919). *A Color Notation*. Boston: Press of GEO H. Ellis CO. https://doi.org/10.1007/s13398-014-0173-7.2

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., … Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*. https://doi.org/10.1038/385432a0

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X12003214

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Computational Biology*, *10*(5). https://doi.org/10.1371/journal.pcbi.1003588

Overgaard, M. (2011). Visual experience and blindsight: A methodological review. *Experimental Brain Research*. https://doi.org/10.1007/s00221-011-2578-2

Park, H. D., & Tallon-Baudry, C. (2014). The neural subjective frame: From bodily signals to perceptual consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2013.0208

Peters, M. A. K., & Lau, H. (2015). Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *ELife*. https://doi.org/10.7554/eLife.09651

Purves, D., Lotto, B., & Polger, T. (2000). Color vision and the four-color-map problem. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/089892900562011

Purves, D., Wojtach, W. T., & Lotto, R. B. (2011). Understanding vision in wholly empirical terms. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1012178108

Railo, H., Koivisto, M., & Revonsuo, A. (2011). Tracking the processes behind conscious perception: A review of event-related potential correlates of visual consciousness.

Consciousness and Cognition. https://doi.org/10.1016/j.concog.2011.03.019

Revonsuo, A.; Johanson, M.; Wedlund, J-E.; Chaplin, J. (2000). The zombies among us. Consciousness and automatic behavior. In A. Rossetti, Y.; Revonsuo (Ed.), *Beyond Dissociation. Interaction between dissociated implicit and explicit processing.* (pp. 331–3351). Amsterdam: John Benjamins.

Revonsuo, A. (2000). Prospects for a Scientific Research Program on Consciousness. In T. Metzinger (Ed.), *Neural Correlates of Consciousness*. Cambridge, MA: MIT Press.

Revonsuo, A. (2006). *Inner Presence: Consciousness as a Biological Phenomenon*. Massachusetts: MIT Press.

Samaha, J., Gosseries, O., & Postle, B. R. (2017). Distinct oscillatory frequencies underlie excitability of human occipital and parietal cortex. *The Journal of Neuroscience*, *37*(11), 2824–2833. https://doi.org/10.1523/JNEUROSCI.3413-16.2017

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience*. https://doi.org/10.1038/nn.2706

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2011.00395

Stoerig, P., & Cowey, A. (1992). Wavelength discrimination in blindsight. *Brain*. https://doi.org/10.1093/brain/115.2.425

Tong, F. (2003). Primary visual cortex and visual awareness. *Nature Reviews*. https://doi.org/10.1038/nrn1055\rnrn1055 [pii]

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*. https://doi.org/10.1016/S0896-6273(00)80592-9

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*. https://doi.org/10.1016/j.tics.2015.10.002

Witzel, C; Gegenfurtner, K. R. (2018). Color perception: Objects, constancy, and categories. *Annual Reviews of Vision Science*, *4*(16), 1–25.

Witzel, C., Valkova, H., Hansen, T., & Gegenfurtner, K. R. (2011). Object knowledge modulates colour appearance. *I-Perception*. https://doi.org/10.1068/i0396

Yang, Z., & Purves, D. (2004). The statistical structure of natural light patterns determines perceived light intensity. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0402192101