Abstract: 148 words
Main text: 7.952 words

Consciousness as a Concrete Physical Phenomenon

Jussi Jylkkä*[1,2,4] & 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

The typical approach to studying consciousness holds that we can only observe the neural correlates of experiences, not the experiences themselves. In this paper we argue, in contrast, that experiences are concrete physical phenomena that can causally interact with other phenomena, including observers. Hence, experiences can be observed and scientifically modelled. We propose that the epistemic gap between an experience and a scientific model of its neural mechanisms stems from the fact that the model is merely a theoretical construct based on observations, and distinct from the concrete phenomenon it models, namely the experience itself. In this sense, there is a gap between any natural phenomenon and its scientific model. On this approach, a neuroscientific theory of the constitutive mechanisms of an experience is literally a model of the subjective experience itself. We argue that this metatheoretical framework provides a solid basis for the empirical study of consciousness.

Keywords: Constitutive mechanisms of consciousness, Neural correlates of consciousness, Epistemic gap, Hard problem of consciousness, Empirical observation, Scientific models, Model-dependent realism, Russellian Monism, Philosophy of science

1. Introduction

The current scientific view about the relationship between consciousness and neural activity is ambivalent. On the one hand, the enterprise to solve the mystery of consciousness using neuroscience is built on the idea that consciousness is identical with neural activity: We know that the brain is necessary for consciousness, and neuroscience has begun to unravel the neural mechanisms that enable subjective, conscious experience (Dehaene & Changeux, 2011; Koch, Massimini, Boly, & Tononi, 2016; Laureys, 2005). On the other hand, scientists and philosophers often assume that science can merely observe *correlations* between neural activity and consciousness, not consciousness per se. After all, how could neural activity be *identical* with consciousness given that the two appear so different? For this reason, neuroscientists search for the neural correlate of consciousness (NCC), which has been defined as the minimal set of neural processes that are together sufficient for a specific conscious experience (Crick and Koch, 1990). Here, we attempt to shed light on this discrepancy.

The present article is an attempt to formulate a plausible and solid metatheoretical foundation for the neuroscientific study of consciousness. This is, in our view, an imperative objective for consciousness science. While consciousness science has established itself as a specific field of empirical science, it seems to be still searching for its identity: Arguably, it is a field that does not yet have a clearly defined explanandum. This is likely to foster false disagreements between theories of consciousness, and contribute to the perception that consciousness science is a less rigorous field than other fields of science (Michel et al., 2018).

Specifically, we wish to address the following: If we assume that an experience is identical with some neural processes, we need to explain why the two *appear* so different, or why there is an epistemic gap between the two. For instance, a person with total congenital achromatopsia arguably cannot know *what it is like* to see the vivid colors of a sunset on a clear evening, even if she perfectly knew all the details of how the human brain works. It appears that science cannot tell us anything about the qualitative or subjective aspects of experiences—the colors of the sunset, the

bitterness of coffee, or the pain of having a manuscript rejected—and thus cannot completely capture their nature (e.g. Chalmers, 1996; Jackson, 1986; Nagel, 1979). 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).

Many attempts in philosophy have been made to explain where the epistemic gap stems from, and the so-called phenomenal concepts strategy is probably the most popular alternative (for a good overview, see Balog, 2012). It aims to reduce the gap into differences between how phenomenal and non-phenomenal concepts refer. Non-phenomenal concepts have a mode of presentation that is distinct from their referent (e.g., we recognize water based on its perceptual features, but not everything that appears to be water is water), but phenomenal concepts are suggested to be more intimately tied with their referent. For instance, it is claimed that phenomenal concepts are recognitional concepts whose mode of presentation is identical with their referent (e.g. Loar, 1990); that phenomenal concepts are similar to indexicals like "here" and "now" (e.g. Ismael, 1999); or that they are quotational concepts, so that e.g. "pain" refers to "that state:___", where the blanks are filled in with pain itself (e.g. Papineau, 2002). Our main focus in this paper is not on differences between how phenomenal and scientific concepts refer, but instead on the question of how, and to what extent, science can model consciousness.

In this paper we take it as our premise that experiences, in all their qualitative richness, are concrete physical phenomena (Strawson, 2008a). To emphasize that empirical explanations of consciousness aim to characterize the causal physical processes that are identical with consciousness—not processes that merely *correlate* with it—we use the concept "constitutive mechanisms of consciousness" (CMC; Revonsuo, 2006). As discussed later in detail, by CMC we mean all those physical processes that together make up consciousness. Similar to biologists who try to describe the hierarchical physical processes that constitute, for example, a living cell or an

amoeba, consciousness science aims to describe the hierarchical constitution of the processes that constitute consciousness. This approach is superficially similar to the aim of describing the NCC. The NCC and CMC approach share the idea that consciousness can be described as a hierarchical process: lower level phenomena (molecules, neurons, their activity) combine to produce a system-level phenomenon (a network of neurons and brain areas), which corresponds to consciousness. The crucial difference between these two views is that the CMC approach implies that the physical phenomenon the scientific theories describe *is* consciousness. The NCC approach, in contrast, suggests that the theories describe a physical phenomenon that merely *correlates* with consciousness. This brings us back to the aim of the present paper: What does it mean to claim that certain neural activation patterns are identical with experiences, given that the two appear so different?

As a summary, we aim to characterize how subjective experiences are related to empirical observations and models about their constitutive mechanisms, and why there appears to be an epistemic gap between the two. We propose that the epistemic gap is distinctness between the scientific CMC-model and the concrete experience. A scientific model is always distinct from the concrete phenomenon it models, and in this sense experiences are not different from other natural phenomena—as suggested by Hawking and Mlodinow (2010), through science we can never know the nature of the world in itself, we can only model it based on observations. Experiences are different from other natural phenomena only because they are the only phenomena in the universe that constitute *our* subjective realm. If we assume that all the properties of experiences are physical and causally efficacious, then they can causally interact with measuring devices. Thus, it is possible to observe and model consciousness itself, not just its correlates. We call our approach *Naturalistic Monism*, because it shares some key components with Russellian Monism (RM), but is completely naturalistic: it does not postulate the existence of properties beyond the scope of natural science.

This article is structured as follows. In the first part, we explain how our approach is motivated in particular by Strawson's (2008a, 2008b) version RM, but also how our view

significantly differs from it. In the second part of the article, we describe in what sense scientific theories are distinct from the phenomena they model, and what implications this has for consciousness research. In the third section, we discuss the implications of Naturalistic Monism from the perspective of science of consciousness, and conclude that neuroscience can study consciousness, not just its correlates. Finally, in the fourth section we briefly discuss the philosophical implications of our approach.

2. Consciousness as a physical phenomenon

Among philosophers, RM has recently gained popularity as an explanation of how experiences are related to their neural mechanisms (e.g. Alter & Nagasawa, 2012; Chalmers, 2016; Goff, 2017; Montero, 2015; Schneider, 2017). It promises to account for why science appears to have trouble in explaining the subjective aspects of experiences, while maintaining that experiences are nevertheless physical phenomena and identical with their neural mechanisms. Naturalistic Monism shares two premises with RM: First, it holds that experiences are concrete physical phenomena; Second, it accepts that science has *in some sense* limited access to their nature. The crucial difference between Naturalistic Monism and RM concerns how the second premise should be interpreted.

Strawson (2008a, 2008b) calls his version of RM "Real Materialism", because it is both materialistic (or physicalistic; Strawson uses the terms interchangeably) and realistic about experiences. He writes:

> "Realistic physicalists, then, grant that experiential phenomena are real concrete phenomena—for nothing in life is more certain—and that experiential phenomena are therefore physical phenomena. It can sound odd at first to use 'physical' to characterize mental phenomena like experiential phenomena, and many philosophers who call themselves materialists or physicalists continue to use the terms of ordinary everyday

> language, that treat the mental and the physical as opposed categories. It is, however, precisely physicalists (real physicalists) who cannot talk this way, for it is, on their own view, exactly like talking about cows and animals as if they were opposed categories. Why? Because every concrete phenomenon is physical, according to them. So all mental (experiential) phenomena are physical phenomena, according to them; just as all cows are animals." (Strawson, 2008b)

Naturalistic Monism endorses this premise, as it is formulated in the above quote. The existence of consciousness is the most certain thing in the world: we can doubt the existence of the whole external world, but not our own experiences. If we are physicalists and take it that experiences feel like something, then we must admit that there exist at least some physical phenomena in the universe that feel like something. However, this is where our agreement with Strawson and other Russellians largely ends. For Strawson continues:

> "I am happy to say, along with many other physicalists, that experience is 'really just neurons firing' […] But when I say these words I mean something completely different from what many physicalists have apparently meant by them. I certainly don't mean that all characteristics of what is going on, in the case of experience, can be described by physics and neurophysiology or any non-revolutionary extensions of them. That idea is crazy. It amounts to radical 'eliminativism' with respect to experience, and it is not a form of real physicalism at all." (ibid)

Strawson's idea is probably the following: if it was possible to characterize the nature of experiences completely in terms of science, there would be no room for phenomenal qualities or what-it-is-likeness, because science appears to say nothing about such properties. Strawson appears to claim that if experiences could be perfectly characterized scientifically, we would be

philosophical zombies (Chalmers, 1996). We strongly disagree with this claim: as we will argue, it conflates experiences as concrete phenomena with their scientific models.

Strawson's view is based on a metaphysical distinction between extrinsic (roughly: relational, structural, and dispositional) and intrinsic (non-relational, non-structural, and non-dispositional) properties (for a useful discussion of the distinction, see Seager, 2006). On this view, science is only concerned with extrinsic properties—how objects behave, which causal dispositions they have, which relations they stand in, what is their structure, and so on. For instance, in Newtonian physics, force is defined in relation to mass and acceleration. For Russellians, this leaves open the question: what is the *non*-relational, *non*-dispositional, and *non*-structural nature of a phenomenon? Russell (1927) himself argued that this intrinsic nature is ontologically "neutral" (that is, neither mental nor non-mental; see Stubenberg, 2016), but Strawson takes it that the intrinsic nature of all physical phenomena is (proto)mental (see also Goff, 2017). To use the current terminology, Strawson would argue that what an experience feels like is part of the intrinsic (science-transcendent) nature of the concrete physical phenomenon, but CMC characterizes the extrinsic (scientifically observable) nature of the same phenomenon.

The intrinsic-extrinsic distinction could be criticized in many ways (see e.g. Hiddleston, 2019; Howell, 2015; Kind, 2015), but here we simply note that the postulation of intrinsic properties is metaphysically promiscuous and unscientific. If we suppose that there is an unobservable and causally impotent intrinsic property corresponding to each observable, causally efficacious property, we double the number of physical properties in the universe solely based on armchair reflection. Moreover, as argued by Ellis (2001), among others, there may be no need to postulate any categorical properties to "ground" causal dispositions or relational properties, because all natural phenomena can be considered as essentially causal-dispositional, or processes. For instance, scientifically it makes no sense to speak of the nature of quarks independently of how they interact with other subatomic particles. According to the standard version of quantum

chromodynamics, it is impossible for quarks to exist in isolation, so it is nomologically impossible for them to possess a non-relational nature.

Because RM implies that experiences are beyond the scope of science, it renders the scientific study of consciousness impossible in principle: science can only study the extrinsic properties of the neural mechanisms of consciousness, not experiences per se. This, coupled with RM's metaphysical heaviness, is probably the reason why RM has not received attention among empirically oriented consciousness researchers. If we could account for the epistemic gap purely epistemologically, without assuming an *ontological* gap between different types of properties, we should do so.[1] Naturalistic Monism aims to do that. It holds that experiences are just like any other physical phenomena in that they can be observed and scientifically modelled. The epistemic gap between experiences and their scientific models does not reflect the existence of any separate types of properties, but instead simply the distinctness between a concrete phenomenon and its scientific model.

3. Distinctness between phenomena and their scientific models

Naturalistic Monism shares with RM the Kantian assumption that science has in some sense limited access to the nature of the phenomena it studies. However, as explained above, whereas RM holds that there exist two distinct types of properties (extrinsic and intrinsic), Naturalistic Monism takes it that there exists only one class of physical phenomena. The limits of science are not due to the existence of ontologically distinct types of properties, but instead to the fact that we can only know the world through observations and scientific models.

Schneider (2017), a proponent of RM, motivates the existence of science-transcendent intrinsic properties by quoting the theoretical physicist Stephen Hawking, who famously asked:

[1] The phenomenal concepts strategy (PCS) aims to do just this: account for the epistemic gap by analyzing the referential mechanisms of phenomenal and scientific concepts (see above). Our approach can be considered as largely compatible with PCS, but we view the question from a different angle: we focus not on semantics, but instead on the nature of science and the relationship between scientific models and the concrete phenomena they model.

> What is it that breathes fire into the equations and makes a universe for them to describe? The usual approach of science of constructing a mathematical model cannot answer the questions of why there should be a universe for the model to describe. (Hawking, 1988, 174).

Strawson (2008b), in turn, quotes the astrophysicist Arthur S. Eddington, who noted that

> [. . .] science has nothing as to the intrinsic nature of the atom. The physical atom is, like everything else in physics, a schedule of pointer readings. The schedule is, we agree, attached to some unknown background. (Eddington, 1929, 259)

But do Hawking and Eddington intend that there would exist some *properties* of physical phenomena completely beyond the scope of science, as Russellians would have it? Or did they simply mean that we cannot *know* the nature of physical phenomena without relying on observation ("pointer readings") or without utilizing scientific models and equations? The former claim is metaphysical, the latter is purely epistemological. At least Hawking intended to make a purely epistemological claim, which he and Mlodinow formulated more explicitly when they introduced the notion of *Model-Dependent Realism* (MDR) (Hawking & Mlodinow, 2010). According to MDR, we can never know the nature of the world as it is in itself, but instead we can only know it through how it affects our senses. Based on observations, we can formulate models of the world in our minds, and we can never say if the models are "true"; we can only evaluate whether they are elegant, can predict observations, or can be used to manipulate phenomena.

Science can model phenomena, but the phenomena themselves are always distinct from their models. According to Naturalistic Monism, an experience is the concrete phenomenon that the scientific model of its constitutive mechanisms describes. For instance, suppose that pain could be

modelled truthfully or adequately[2] as T-type interaction between neural modules M1 and M2 (or T-interaction, for short). The central point here is that "T-interaction" is just pain's scientific model, formed in the mind of a scientist based on observations such as which neural events are correlated with reports of pain, avoidance behavior, certain facial expressions, and so on. Even a scientific realist, who assumes that scientific theories truthfully capture the nature of natural phenomena, cannot hold that pain is *literally* nothing but "T-interaction", because "T-interaction" is a theoretical *model* of pain, and distinct from pain itself. Instead, the realist should be interpreted as claiming that pain can be truthfully *modelled* as "T-interaction". This is compatible with the fact that when the concrete process scientifically modelled as "T-interaction" happens in a subject, the occurrence of the process feels like something for her—the scientific realist would simply say that the feel of pain can be truthfully modelled as "T-interaction".

Because any concrete phenomenon is always distinct from its scientific model, a neuroscientific CMC_E-model of an experience E is never identical with E; it only refers to E. Whereas science can only model experiences through observation, we know our own experiences immediately. My pain is not something outside of me that I observe, it is part of my subjective realm. It constitutes the physical process that is my consciousness, which could be observed and modelled by a scientist as CMC_{pain}. To paraphrase Edelman and Tononi (2000), "Unlike any other entity, […] with consciousness *we are what we describe* scientifically" (p. 14, italics in the original). The reason why we easily think that there is an epistemic gap just in the case of experiences is because our experiences constitute us, and thereby we have non-empirical access to their nature—access that is not mediated through observations and models. Naturalistic Monism implies that there is an epistemic gap between *any* concrete phenomenon and its scientific model, due to the distinctness between the model and the phenomenon. As discussed later in the article (section 5.1.), how wide we take the gap to be depends on which philosophy of science we endorse.

[2] Whether we take a scientific model to be capable of truthfully modelling a phenomenon depends on which philosophy of science we endorse; see section 5.1.

Naturalistic Monism implies that what an experience feels like is part of its nature in itself, independently of observations, scientific models, or other theoretical characterizations. In fact, already describing an experience as "feeling like something" is already a conceptualization, and it could be debated whether it adequately or truthfully captures the nature of the experience. Descriptions of consciousness typically depend on the person's favorite philosophical theory, religion, culture, or level of education. Consciousness could be said to be "neural activation", "qualitative", "phenomenal", "feeling like something", "non-relational", "non-structural", consisting of "qualias", "soul", "Brahman", and so on. All such *definitions* of experiences could be questioned, but that does not amount to questioning the nature of experiences themselves. If you take René who believes in an immaterial soul, David who believes in qualia, and Daniel who denies the existence of qualia, and stick them with a needle, the same kind of natural phenomenon is instantiated in all of them. The concrete phenomenon that occurs in them is of the same type, even if they define it in very different ways—nature does not care about our definitions.[3] What pain "feels like" for a subject S is simply the happening of pain in S, the occurrence of the natural phenomenon. When pain happens in me, I can minimally say it feels like "this", referring to the happening of my pain experience (Ismael, 1999; Papineau, 2002). Consciousness science aims to model such processes, and can in this sense model what experiences feel like. In the form of an argument:

1) What pain feels like for S = the happening of pain in S.

2) The happening of pain in S is a concrete physical process.

3) Science can model all concrete physical processes.

4) Thus, science can model what pain feels like.

[3] This is a simplification, because conceptualizing an experience can modify it—for instance, verbalizing an emotion can weaken its intensity (e.g. Torre & Lieberman, 2018).

It is common to suppose that knowing a scientific model of pain does not give knowledge of what pain feels like. In contrast, according to Naturalistic Monism, knowing the pain-model *does* give knowledge of what pain feels like, but only knowledge in terms of a scientific model, which is distinct from concrete pain. What concrete pain feels like is simply the happening of pain in us, a process that science aims to model with the pain-model. We can say that pain is nothing but "T-interaction", but this does not amount to reducing concrete pain into the theoretical T-interaction-model. Instead, pain, including what we call its "what-it-is-likeness", is the concrete phenomenon that science *models as* T-interaction. The take-home message is that we need not postulate the existence of any properties ("qualia") for the what-it-is-likeness of pain over and above the properties described by the scientific pain-model. Once we have a scientific model of pain that is perfectly isomorphic with pain as experienced, we have a scientific model of pain's phenomenology. For instance, the experienced intensity or sharpness of pain can be modelled as parameters in the pain-model that correspond to such qualities—any change in the experienced sharpness of pain corresponds to a change in the corresponding parameter (see section 4.1.). However, this does not mean that pain itself would be just the theoretical parameters; it is what the parameters model.[4]

4. Studying consciousness scientifically

We have argued, first, that consciousness should be viewed as a concrete physical process in the world in the same sense as lightning, photosynthesis, and life—experiences are of the same ontological type as all physical phenomena, there is only difference in complexity. Second, we have stressed that scientific models are always distinct from the phenomena they model. Because experiences are concrete phenomena, they can causally interact with other physical phenomena,

[4] Dennet (1991) makes a similar case, but he does not clearly distinguish between concrete phenomena and their theoretical models. Arguably, this has led people to accuse him of being an eliminativist about experiences. However, it is one thing to be an eliminativist about "qualia", a theoretical concept which holds that experiences involve completely non-structural, non-relational, and non-dispositional properties, and a completely different thing to be an eliminativist about concrete experiences. Dennett is an eliminativist about the former, but not about the latter. On our view, he would simply claim that experiences can be *modelled* completely functionalistically, but this does not amount to claiming that the modelled phenomena themselves would not feel like something (in the everyday sense of "feeling like something").

including observers and their measuring devices. Science can model experiences as CMCs. For a neuroscientist, the experience of a subject is the "external" phenomenon that affects her measuring devices, and which she aims to describe with the CMC-model. From the subject's perspective, her experience is an "internal" phenomenon, part of her subjective realm.

We will next illustrate Naturalistic Monism with the help of a neuroscientific example. Consider a visual experience of a cat (Figure 1). On our approach, a subjective cat experience is the physical process in the world that underlies neuroscientific observations that we interpret in terms of the CMC_{cat}-model (a model of the constitutive mechanisms of a cat experience). Thus, the terms "cat experience" and "CMC_{cat}" refer to the same concrete phenomenon.[5] Whereas the term "experience of cat" is typically used to refer to an experience when it is part of a subject's consciousness, the neuroscientific term "CMC_{cat}" is used in scientific contexts, where it refers to the experience as the external phenomenon that causes scientific observations. Experiences are complex physical processes that are composed of lower-level phenomena, and the composition can be described scientifically. Science can describe how experiences are composed of lower level processes in the same way as it explains how hydrogen and oxygen form water. This framework is summarized in Figure 1.

[5] Here it is important to distinguish between experience- and CMC-*types*, and their *tokens*, such as a particular cat-experience. We do not wish to take a stand on this question here, but we assume that at least each experience token is identical with what its complete neuroscientific model token refers to. Types are more complex because they are always abstractions; we could question whether there can exist any cat-experience type. Models, in turn, are themselves abstractions based on observations of tokens, and cannot be considered as strictly identical with any particular experiences. For a discussion on the type-token distinction in physicalism, see e.g. Stoljar (2001).

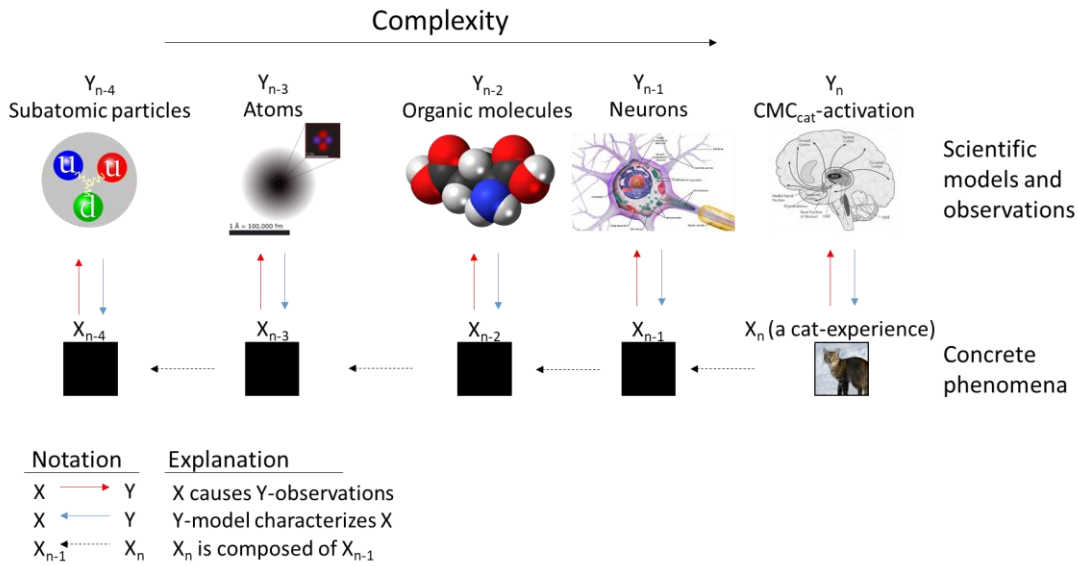

Figure 1. *A schematic presentation of Naturalistic Monism. Empirical observations (upper row) are caused by concrete physical phenomena (red arrows), and empirical models characterize the concrete physical phenomena (blue arrows). Complex physical phenomena are composed of more basic level phenomena (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 models. Because experiences are complex physical phenomena, they are composed of lower-level constituent processes. The lower-level constituents of experiences are shown in black because they are not part of our consciousness; they can only be characterized empirically. (All images under Creative Commons license.)*

The cat-experience is depicted in Figure 1 as a complex physical process that is composed of lower-level constituent processes. This can be empirically modelled by defining how the CMC_{cat}-process is related to its constituent processes. It is often argued that consciousness cannot be fully explained by analyzing the interactions between its lower-level constituents (this is the so-called "combination problem" that RM faces; see Chalmers, 2017). This argument conflates concrete phenomena with their scientific models: one can never "reduce" *any* concrete phenomenon to lower level processes, because reduction is always explanatory. To give an example, one can *explain* how the functioning of an amoeba is based on the interactions between the molecules that constitute it, and in this sense reduce the *scientific model* of the living amoeba to *models* of lower-level phenomena, such as interactions between proteins, enzymes, and so on. This does not mean that the actual, concrete amoeba is "reduced" to anything else; the concrete amoeba *is* the complex system composed of the interaction between molecules, enzymes, and so on. In the same way, one can *model* how an experience is based on lower level processes, but this does not amount to "reducing"

the concrete experience into anything else. When we explicate how a cat-experience is composed of lower level processes (the lower axis in Figure 1), we do it in terms of describing how the CMC_{cat} is composed of lower-level processes, such as the activation of individual neurons (the upper axis in Figure 1). Crucially, when we scientifically reduce an experience into lower-level processes, we do so only descriptively. "Reducing" the concrete experience—literally tearing it to parts—would amount to destroying it. The crucial point here is that if we accept that scientific models of experiences are literally models of experiences, then it is possible to scientifically model how subjective experiences are based on lower-level processes.

The black boxes in Figure 1 might appear to motivate some kind of mysticism about the nature of (non-mental) physical phenomena, similar to Strawson's "intrinsic nature" that cannot be empirically captured. However, this mysticism is inherent to all science: we can never know the nature of the world as it is in itself, we can only model it based on how it affects our senses. In Figure 1, the lower level boxes are mostly black to emphasize that the phenomena exist independently and outside of how we scientifically model them, but this does not imply that they are *completely* epistemologically opaque to us: we *can* know their nature, but *only* through observations and models. Moreover, it is important to notice that whether a box is black or not depends on perspective: my consciousness is not a black box for me, but it is for you, because your knowledge of my consciousness is based solely on observations and models (your model of my mind).

4.1. Towards an empirical model of consciousness

Because Naturalistic Monism implies that an experience is a concrete physical phenomenon, and that all the properties of experiences are physical properties, it must be possible to observe and scientifically model all the characteristics of an experience. Hence, based on the current framework, empirical science can study consciousness, not just its correlates. This is an important upshot, as it collapses the problem of consciousness into a standard problem of science.

Specifically, we suggest that a scientific model of consciousness should describe the following (Figure 2):

1. *Constituents*. Because consciousness is a complex physical process, science should describe its constituents. This explanation describes what consciousness is "made of".
2. *Isomorphy*. A complete model of consciousness should accurately describe all the details about the contents of consciousness, including its phenomenology.
3. *Etiology*. Understanding the etiological basis of consciousness, from immediately preceding causes to its phylogenetic basis, is crucial for understanding phenomenology.
4. *Causal power*. Being a concrete physical phenomenon, consciousness can causally interact with other physical processes inside the brain. A key part of explaining consciousness is explaining what it can do.

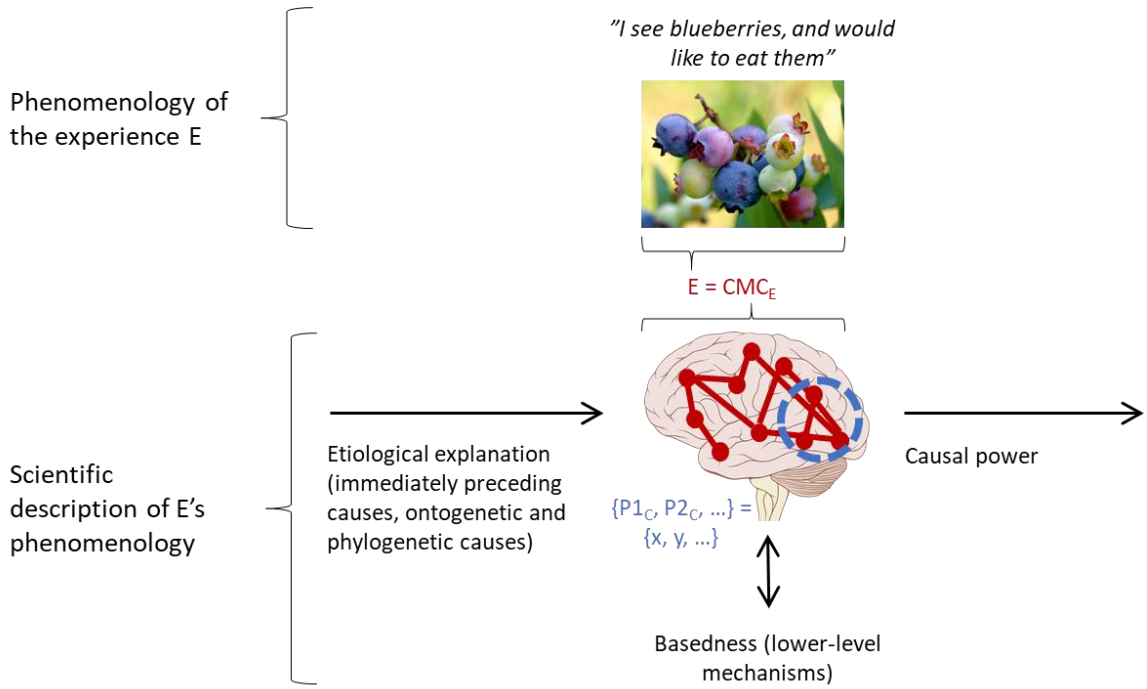

Figure 2. *Understanding phenomenology empirically. Phenomenology of experience E is depicted in the figure by the picture of blueberries. We assume that at this moment the subject experiences seeing blueberries, and feels an urge to eat them. The scientific model of E is here depicted by the image of 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 CMC$_E$ is a model of the experience E. Hence, CMC$_E$ and E, although appearing to be different, are isomorphic, and E can be described by a detailed empirical model (CMC$_E$). To understand the "mechanism of consciousness", science must describe the lower-level processes*

that constitute it, and describe the preceding causes of E (etiology; this include e.g. phylogeny). 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.]

4.1.1. Lower-level 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. The constituents are for CMC what hydrogen and oxygen are for water: hydrogen and oxygen do not cause water, they constitute it. The constituents of consciousness are often called the "mechanism of consciousness": Which neural processes determine whether an individual is conscious or unconscious? Which neural processes determine whether a visual stimulus will be consciously perceived? Examples of theoretical attempts to answer such question include the global workspace model (GWT; Baars, 1988; Dehaene & Changeux, 2011), integrated information theory of consciousness (IIT; Oizumi, Albantakis, & Tononi, 2014), and higher-order thought theory of consciousness (Lau & Rosenthal, 2011). These theories aim to describe how consciousness "emerges" from neural activity. According to Naturalistic Monism, these theories aim to describe the CMC. Crucially, because the CMC is a model of consciousness, explaining the constituents of CMC amounts to explaining the constituents of consciousness itself—not just something that correlates with consciousness.

Consider the GWT, for example. According to the theory, consciousness is a process where a large number of individual modules share their information with each other to enable flexible behavior and maintenance of information (Baars, 1988). According to the neural model of the theory, this corresponds to synchronous interactions between specific neural populations in frontal, parietal and sensory cortical areas (Dehaene & Changeux, 2011). These individual processes are not conscious alone, but become conscious when they interact in the way that the theory describes. These constituent processes are 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.

GWT does not make claims about how the subjective aspects of consciousness (phenomenology) are enabled by activation of the global workspace, but according to Naturalistic Monism, the concrete phenomenon that GWT models *is* consciousness and has a certain phenomenology for the subject in whom it happens (causal power of consciousness is discussed further in section 4.1.4.).

4.1.2. Isomorphism

Because the phenomenology of an experience is a concrete physical process, any change in phenomenology can in principle be observed and modelled by a specific change in the model. Thus, a complete CMC_E- model of an experience E must be perfectly isomorphic with the experience: there must be a one-to-one correspondence between the empirically observed physical phenomenon (CMC_E) and the subjective experience E (Fingelkurts & Fingelkurts, 2004; Revonsuo, 2000). This means that an empirical theory of consciousness, if sufficiently detailed, can provide a complete scientific description of an experience and its phenomenology. To do this, the theory must be able to define which empirically observed physical processes correspond to different contents of experiences, and how these contents combine to form complex experiences. For instance, from the perspective of conscious vision, the requirement of isomorphy means that the model must predict the phenomenology of vision from the panoramic, egocentric perspective – not from the gaze-centered, retinotopic frame (Land, 2012). According to the integrated information theory of consciousness (Oizumi et al., 2014) the structure of recurrent networks ("qualia space", to use IIT's terminology) defines phenomenology. According to Naturalistic Monism, such models are descriptions of the concrete physical phenomenon that is consciousness.

The reader may, at this point, insist that isomorphic descriptions of CMCs would not explain the "qualitative feelings" of experiences, or answer why e.g. blue feels like blue, and hence the

description would not be an explanation of what blue feels like. Such counter-arguments are ill-posed, because they conflate concrete phenomena with their scientific models. What blue feels like for S is the happening of CMC_{blue} in S, a concrete process that can be perfectly modelled scientifically—but just *modelled*. Grasping the CMC_{blue}-model affords knowledge of how an experience of blue can be scientifically *described*, whereas what blue feels like is just the happening of the concrete phenomenon, which is distinct from the scientific model. 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. From the perspective of Naturalistic Monism, 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.

4.1.3. Etiology

In trying to explain what must happen in the brain so that, for instance, visual information crosses the threshold to consciousness, neuroscience examines the causal chain of events that lead to it (Dehaene & Changeux, 2011; Railo, Koivisto, & Revonsuo, 2011). In other words, they describe the etiology of conscious vision. According to a widely held view, conscious vision requires, first, that the visual stimulus evokes sufficiently strong activation in different areas of the brain. The processes that define whether or not the activity related to a visual stimulus crosses the threshold to consciousness happen after these "prerequisite" processes (Aru, Bachmann, Singer, & Melloni, 2012). One of the challenges of empirical science is to characterize this etiological chain of events that leads to conscious vision, that is, the formation of the CMC that corresponds to experience of that stimulus.

In addition to explaining the immediately preceding causes that lead to consciousness, science can describe longer time scales such as evolution or individual development (see Revonsuo, 2006). For example, a complete 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. Similar argument, which attempts to account visual phenomenology in wholly empirical terms by describing how frequently different stimulus combinations have been encountered during the evolution (Purves, Wojtach, & Lotto, 2011). This view is built on the assumption that consciousness is not an epiphenomenon (e.g. Jackson, 1982), but instead that it has power to influence other physical processes.

4.1.4. Causal power

The idea that experiences are something different from other physical phenomena is deep-rooted in philosophy and science. For instance, Edelman (2003), who builds an elegant theory of consciousness on a bio-physical foundation, argues that conscious states are "informational even if not causal" (p. 5523). This way, consciousness easily appears epiphenomenal, which is strongly at odds with our subjective experience as well as the empirical scientific assumption that human consciousness is the product of evolution and plays a causal role in behavior. 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 (Block, 1995; Carter et al., 2018; Dehaene & Changeux, 2011). 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. One of the goals of empirical science of consciousness is to explain what gives conscious processes their power to potentially elicit fast and accurate actions, and flexible voluntary behavior[6]. 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 (cf. the example about pain, in previous subsection).

5. Philosophical implications

In this section, our aim is to briefly address the main philosophical implications of Naturalistic Monism. The discussion is not intended as exhaustive, but rather as pointing towards future research.

5.1. Philosophies of science and consciousness

If we consider consciousness as a concrete physical phenomenon, then the question about the relationship between an experience and its mature scientific model comes at least partly down to which philosophy of science we endorse (in addition to the details of the model itself). A scientific model is always distinct from the concrete phenomenon it models, and there is always a gap between the two in this sense. Whether we take there to be an *epistemic* gap, and how wide we take it to be, depends on whether we endorse realism or antirealism about science. Roughly, according to realism, mature scientific theories can be true, and truthfully describe the nature of the world (Psillos, 1999). Thus, the epistemic gap between a phenomenon and its mature scientific model is non-existent or narrow. According to antirealism, on the other hand, we always have to remain

[6] While stimuli that are not consciously experienced may influence individual's behavior (Cowey, 2010; Hannula, Simons, & Cohen, 2005), 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). Moreover, recent research has also shown that early studies overestimated the power of unconscious processes to influence behavior (Ffytche & Zeki, 2011; Newell & Shanks, 2014; Overgaard, 2011; Peters & Lau, 2015)

agnostic about the truth of scientific models. For instance, according to constructive empiricism (van Fraassen, 1980), scientific theories can be considered as *empirically adequate* if they are isomorphic with observations, but we can never say whether they are literally true. We could say that, on this view, there is always an indefinitely wide epistemic gap between a phenomenon and its model, or that it makes no sense to speak of such a gap.

Scientific realism implies that if we said that our experiences involve aspects that are completely qualitative (not structural-relational), we would be mistaken: experiences cannot involve such properties, because science does not postulate them (we assume that all scientific models are structural-relational). On this view, what we call a "qualitative feeling" is really a structural-relational process. Importantly, this approach does not deny that experiences feel like something; it only holds that if we *characterized* the feel as something purely qualitative, we would be mistaken. Terms like "qualitative" or "feeling" are only concepts which refer to certain concrete processes that happen in us. The processes are the way they are, no matter how we refer to them. Scientific realists would hold that experiences can be truthfully modelled completely in structural-relational terms, but this does not, and cannot, amount to eliminating any aspects of the experiences themselves. It may be difficult to see how, say, the redness of red could be a structural-relational process, given that introspection about redness does not reveal that it has such nature. However, we are often mistaken about the nature of our experiences, as demonstrated by many illusions, such as the checker shadow illusion. Moreover, from the fact that a model M applies to an experience E, it does not follow that a subject experiencing E would introspectively see that M applies to it (cf. Dennet, 1991).

If we endorse antirealism, such as van Fraassen's empiricism, we can only say whether a neuroscientific theory of consciousness is compatible with observations, not whether it "captures the nature" of what it models. It could be said that, for the antirealist, there is *always* an epistemic gap between *any* concrete phenomenon and its scientific model, and that experiences are not different in this respect. It is noteworthy that Naturalistic Monism is not committed to either realism

or antirealism; this is an independent question. Naturalistic Monism implies that there is always *distinctness* between a phenomenon and its model, but not necessarily an *epistemic* gap: if we endorse scientific realism about neuroscientific theories of consciousness, there is only distinctness between an experience and its model, but not an epistemic gap.

Detailed discussion of how consciousness is related to different philosophies of science is beyond the scope of this paper. The main point is that if we consider experiences as being of the same (ontological) type as any other physical phenomenon, then the question about the relationship between an experience and its scientific model boils down to the general question about the relationship between *any* phenomenon and its model.

5.2. Panpsychism

According to Naturalistic Monism, experiences are concrete physical phenomena that are based on lower-level physical processes, and thereby ontologically continuous with them. Does it follow that *all* physical phenomena are of an "experiential nature", because they are of the same ontological type as experiences (cf. the black boxes in Figure 1)? In our view, this is not a necessary conclusion. An experience E can be scientifically modelled as CMC_E, a system-level phenomenon in the brain that is based on the activity of individual neurons. In this sense, consciousness is continuous with other physical phenomena, all the way down to subatomic particles. However, insofar as individual neurons (not to speak of quarks) alone do not realize the CMC_E-process, they are not conscious, at least in the same sense as we are. On this line of argumentation, it is possible for consciousness to emerge from something that is not conscious, and it is the task of empirical theories of consciousness to describe exactly how this happens.

On the other hand, we have no non-empirical or non-model-mediated access to the nature of other physical phenomena besides our experiences. We can extrapolate from our own case and infer that at least cats and cows are likely to have similar experiences as us, given that they show similar behaviors and have similar neuroanatomy. It is more difficult to say anything about the possible

inner lives of insects, amoebas, or quarks. This difficulty, however, stems from the limits of science, not from consciousness being a special type of phenomenon.

5.3. *Zombies, Mary, and the "hard problem"*

Zombies (Chalmers, 1996) are creatures that appear just like normal human beings in every functional and observable respect, but for whom "the lights are out": it does not feel like anything to be a zombie. When we imagine zombie-pain, we imagine something that would fit our scientific model of pain, but which would not be real pain. If such fake-pains are conceivable, then arguably we could conceive of any fake natural phenomena. For instance, we can imagine that we live in a massive computer simulation, where our observations of electrons are not caused by real electrons, but instead by computer algorithms. If such imaginings are possible at all, what they teach us is that our scientific models and observations are always distinct from the concrete phenomena that underlie them. It does not tell us anything about the nature of experiences or electrons, except for that they are concrete phenomena and independent of their scientific models and observations.

The same reasoning can be applied to the Mary case, a neuroscientist who has never experienced colors (Jackson, 1986). Mary knows everything about the visual color system in the brain, but she does not know what it is like to see red. The conclusion that the thought experiment is supposed to demonstrate is that there are some facts about experiences that cannot be described by scientific models. But the conclusion is false: Mary *does* know what it is like to see red, but only in terms of a scientific model. What it is like for S to see red is the happening of a red experience in S. Because science can model such processes, it can thereby model what it is like to see red—but only *model*. When Mary sees red for the first time, the concrete phenomenon whose model she already knew *happens* in her. What the Mary case teaches us is that a concrete phenomenon can be quite different from what a scientific model says it is. But this does not imply that there would be some properties ("qualia") that cannot be scientifically modelled, it only shows that a concrete phenomenon is distinct from its model.

6. Conclusions

We have argued that once we clearly distinguish between concrete phenomena and their scientific models, the epistemic gap – why our experiences appear so different from neural activity – dissolves into distinctness between the model and the phenomenon. From the perspective of Naturalistic Monism, science can observe and model consciousness, not just their correlates. We hope that our proposal helps to establish a solid ground for the empirical science of consciousness, as it collapses the problem of consciousness into a standard problem of science. Scientific explanation never affords us knowledge of the nature of phenomena in themselves. However, if we accept that our subjective experiences are concrete physical phenomena, then science can provide a complete and exhaustive *model* of experiences, including what they "feel like".

7. Acknowledgements

J.J. was supported by the Abo Akademi University Endowment (the BrainTrain project and a 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.

10. References

Alter, T., & Nagasawa. (2012). What is Russellian Monism. *Journal of Consciousness Studies*, *19*(9–10).

Aru, J., Bachmann, T., Singer, W., & Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2011.12.003

Baars BJ, & Baars, B. J. B. (1988). *A cognitive theory of consciousness*. *A Cognitive Theory of Consciousness*. https://doi.org/10.1108/09513551111163639

Balog, K. (2012). In Defense of the Phenomenal Concept Strategy1. *Philosophy and*

Phenomenological Research, *84*(1), 1–23. https://doi.org/10.1111/j.1933-1592.2011.00541.x

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.

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. (1996). *The Conscious Mind: in Search of a Fundamental Theory*. Oxford: Oxford UP.

Chalmers, D. (2016). Panpsychism and panprotopsychism. In G. Büntrup & L. Jaskolla (Eds.), *Panpsychism: Contemporary Perspectives*. Oxford: Oxford UP.

Chalmers, D. (2017). The combination problem for panpsychism. In *Panpsychism. Contemporary Perspectives*. Oxford: Oxford UP.

Cowey, A. (2010). The blindsight saga. *Experimental Brain Research*. https://doi.org/10.1007/s00221-009-1914-2

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

Eddington, A. (1929). *The Nature of the Physical World*. Cambridge: Cambridge UP.

Edelman, G. M. (2003). Naturalizing consciousness: A theoretical framework. *Proceedings of the*

National Academy of Sciences. https://doi.org/10.1073/pnas.0931349100

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

Ellis, B. (2001). *Scientific Essentialism*. Cambridge: Cambridge UP.

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

Goff, P. (2017). *Consciousness and Fundamental Reality*. Oxford: Oxford UP.

Hannula, D. E., Simons, D. J., & Cohen, N. J. (2005). Imaging implicit perception: promise and pitfalls. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn1630

Hawking, S. (1988). *A Brief History of Time*. New York: Bantam Books.

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

Hiddleston, E. (2019). Dispositional and categorical properties, and Russellian Monism. *Philosophical Studies*, *176*(1), 65–92. https://doi.org/10.1007/s11098-017-1006-2

Howell, R. (2015). The Russellian Monist's Problems with Mental Causation. *The Philosophical Quarterly*, *65*(258), 22–39. https://doi.org/10.1093/pq/pqu058

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.

Jackson, F. (1986). What Mary Didn't Know. *The Journal of Philosophy*, *83*(5), 291–295. https://doi.org/10.2307/2026143

Kind, A. (2015). Pessimism about Russellian monism. In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism* (pp. 401–421).

Oxford: Oxford UP.

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

Land, M. F. (2012). The operation of the visual system in relation to action. *Current Biology*. https://doi.org/10.1016/j.cub.2012.06.049

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

Laureys, S. (2005). The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2005.10.010

Loar, B. (1990). Phenomenal states. *Philosophical Perspectives*, *4*, 81–108.

Michel, M., Fleming, S. M., Lau, H., Lee, A. L. F., Martinez-Conde, S., Passingham, R. E., … Liu, K. (2018). An informal internet survey on the current state of consciousness science. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2018.02134

Montero, B. (2015). Russellian physicalism. In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the Physical World* (pp. 209–223). Oxford: Oxford UP.

Nagel, T. (1979). What it is like to be a bat? In *Mortal Questions* (pp. 165–180). Cambridge: Cambridge UP.

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

Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford UP.

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

Psillos, S. (1999). *Scientific Realism: How science tracks truth*. London: Routledge.

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.

Russell, B. (1927). *The Analysis of Matter*. London: Kegan Paul.

Schneider, S. (2017). What breathes fire into the equations? *Journal of Consciousness Studies*, *24*(9–10).

Seager, W. (2006). The “intrinsic nature” argument for panpsychism. *Journal of Consciousness Studies*, *13*(10–11), 129–145.

Stoljar, D. (2001). Physicalism. In *Stanford Encyclopedia of Philosophy*. Retrieved from https://plato.stanford.edu/entries/physicalism/

Strawson, G. (2008a). Real materialism. In *Real Materialism and Other Essays* (pp. 19–52). New York: Oxford UP.

Strawson, G. (2008b). Realistic monism: Why physicalism entails panpsychism. In *Real*

Materialism and Other Essays (pp. 53–74). New York: Oxford UP.

Stubenberg, L. (2016). Neutral Monism. In *Stanford Encyclopedia of Philosophy*. Retrieved from https://plato.stanford.edu/entries/neutral-monism/

Torre, J. B., & Lieberman, M. D. (2018). Putting feelings into words: Affect labeling as implicit emotion regulation. *Emotion Review*, *10*(2), 116–124.

van Fraassen, B. (1980). *The Scientific Image*. Oxford: Oxford UP.