8-2024

The Hard Problem of Consciousness and Neurobiological Naturalism

Bailey Villarreal

doi: https://doi.org/10.57709/37124616

The Hard Problem of Consciousness and Neurobiological Naturalism

by

Bailey D. Villarreal

Under the Direction of Dr. Sarah F. Brosnan, PhD

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2024

ABSTRACT

In this thesis, I present Feinberg and Mallatt's (2020) theory of neurobiological naturalism and attempt to situate it within the broader philosophical discussion of consciousness. They propose what they consider to be the neurobiological components sufficient for phenomenal consciousness to have evolved and claim to have filled the explanatory gap in our understanding of how the brain gives rise to experience. I will argue that their claim for doing so is premature and rests on a fundamental misunderstanding of what distinguishes the hard and easy problems of consciousness, which results in a vicious circularity. I explore how Feinberg and Mallatt could respond and remain internally consistent with their theory. One way is an appeal to what I have called *virtual irreducibility*, which is an epistemic claim regarding the limitations of our knowledge of weakly emergent system properties. I consider this response, among others, and demonstrate why they are problematic.

INDEX WORDS: Consciousness, Evolution, Emergence

The Hard Problem of Consciousness and Neurobiological Naturalism

by

Bailey D. Villarreal

Committee Chair: Sarah F. Brosnan

Committee: Andrea Scarantino

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2024

DEDICATION

I dedicate this thesis to my family, friends, mentors, and so many more. All of whom, I could not have done without.

ACKNOWLEDGEMENTS

I would like to acknowledge the expertise and advice of my thesis advisor Dr. Sarah F. Brosnan who has been nothing but the greatest help throughout this endeavor. I would also like to offer special thanks to my reader Dr. Andrea Scarantino for all of his helpful feedback on multiple drafts of this manuscript. Lastly, I would like to acknowledge all of the support I have received over the years from loved ones who have stood by and listened to my ideas.

TABLE OF CONTENTS

1 INTRODUCTION

In this thesis, I will present Feinberg and Mallatt's (2020) theory of neurobiological naturalism and attempt to situate it within the broader philosophical discussion of consciousness. There are different ways consciousness has been defined throughout the literature. However, for this thesis, phenomenal consciousness will be the primary focus, and I will sometimes use the terms consciousness and phenomenal consciousness interchangeably. *Phenomenal consciousness* is the phrase used to refer to the subjective, first-person qualitative aspects of experience. Philosophers and scientists alike have struggled to solve what has been called the hard problem of consciousness, which is the question of how the physical composition of the brain can give rise to experience. Philosophers like Levine (1983) have called this the explanatory gap, and the challenge of closing it represents the problem. Philosophers like Chalmers (1995) have distinguished the hard problem of consciousness from some easy problems which appear more accessible to methods of investigation. Easy problems, according to Chalmers (1995), include aspects of attention, the ability to categorize things and respond to environmental stimuli, reportability of mental states, and some others (p. 2). These problems can be studied using standard investigatory practices within cognitive science and require a level of explanation specifying their various functions (Chalmers, 1995, p.2).

In contrast, the nature of *qualia*, the fact that sensory experiences feel a certain way as opposed to some other way, represents one of the hard problems insofar as standard practices fail to explain these qualitative differences. For example, objectively observable neural activity does not disclose the redness or sweetness of an apple (Feinberg, 2011, p. 19). In their account, Feinberg and Mallatt claim that phenomenal consciousness is a natural product of evolution within certain animal lineages. They delineate at length what they consider to be the neurobiological components sufficient for phenomenal consciousness to have evolved and their functional characteristics. They claim to have filled the explanatory gap in our understanding of how the brain gives rise to experience. I will argue that this claim for doing so is premature and rests on a fundamental misunderstanding of what distinguishes the hard and easy problems of consciousness, resulting in a vicious circularity. Feinberg and Mallatt could respond by appealing to what I have called *virtual irreducibility*, which is an epistemic claim regarding whether we can have full knowledge of something like phenomenal consciousness or only to a very limited extent. I consider this response and demonstrate why it is problematic. Feinberg and Mallatt's work on the evolution of consciousness represents a valuable contribution to both scientific and philosophical disciplines. As such, this thesis aims to summarize their work while elucidating some of the scientific and philosophical challenges evolutionary studies of consciousness face. I will begin by presenting philosophical treatments of consciousness before situating Feinberg and Mallatt's work within the context of representational theories, which are quite similar to their own views. Then, I will introduce their theory of *neurobiological naturalism* before moving on to a philosophical critique of what I have called their *special features argument* for having filled the explanatory gap in our understanding of phenomenal consciousness.

2 PHILOSOPHICAL TREATMENTS OF CONSCIOUSNESS

In this section, I will introduce some of the core features of the philosophical debate concerning consciousness. I will also present a brief overview of representational theories of consciousness because they are most similar to the theories held by the primary authors I will be engaging with. This section will also cover how representational theories of consciousness can contribute to evolutionary understandings of consciousness.

2.1 Phenomenal Consciousness

According to Nagel (1974), whatever form it takes, if an organism is conscious, there must be something that it is like to be that organism (p. 436). This definition of consciousness, often called phenomenal consciousness, has received popular support within the philosophical and scientific literature on consciousness. Nonetheless, consciousness has been defined in many other ways, and there is no clear consensus on how it operates or evolved. It may very well be an undeniable aspect of our existence, yet it defies simplification and our best attempts at understanding even its most basic qualities. Philosophers of consciousness have, however, widely agreed on some notable distinctions. Namely, we can and should distinguish *creature consciousness* from *mental-state* consciousness (Carruthers, 2000, pp. 254-255). According to Carruthers (2000), we can ascribe consciousness to any person or organism, but to do so is quite different from claiming that their mental states are conscious.

Moreover, he says that we can distinguish *intransitive* and *transitive* variants of creature consciousness. Intransitivity is to say of any particular creature that it is conscious *simpliciter*, meaning that it is in a wakeful state instead of asleep. Transitivity implies that if a creature is conscious of Ψ, it perceives Ψ (Carruthers, 2000, pp. 254-255). In other words, if a creature is *transitively* conscious, it is in a wakeful state and has sensory experiences of the world. On the subject of *mental-state consciousness*, Carruthers (2000) remarks that a major distinction is often drawn between phenomenal consciousness (p-consciousness) and other *functionally definable* variants like Block's (1995) *access consciousness* (Carruthers, p. 255).[1] According to Block (1995), a mental state demonstrates access consciousness (a-consciousness) if a representation of its content is "(1) …poised for use in reasoning, (2) poised for rational control of action, and (3) poised for rational control of speech" (p. 231).[2] Furthermore, he says that a-conscious states can only ever be transitive in the sense that wakefulness is a necessary condition for the modification of behaviors by representational contents of any sort (Block, 1995, p. 232). For a creature's behaviors to be modified by mental states whose representational contents originate from the external environment, they need to be in a wakeful state, actively engaging with that environment. Alternatively, p-conscious states can sometimes be intransitive. During instances of dreaming, for example, there is undoubtedly a phenomenal character to the experience, but in the absence of wakefulness (Block, 1995, pp. 232-233). However, there is still considerable disagreement over whether mental states can be considered p-conscious without consciousness of any functionally definable variety (Carruthers, 2000, p. 255). Decidedly less controversial is the notion that p-consciousness is the most philosophically problematic (Carruthers, 2000, p. 255). In a thought experiment first proposed by Nagel (1974), he asked us to imagine what it is like to be a bat. This exercise was meant to elucidate our limited capacity to imagine sensory experiences, like echolocation, vastly different from our own. Moreover, what one can imagine only amounts to imagining themselves as bats, without revealing anything bats

[1] If an aspect of cognition is functionally definable, it can be described in terms of what it does (Block, 1995, p. 232).

[2] What *representations* are and how to define them and their respective *contents* is an ongoing debate within the cognitive sciences and philosophy of mind, which I cannot engage in here. However, for the purposes of this paper, it will suffice to consider them to be information-bearing structures within the brain (see Pitt, 2022).

experience, presenting an intractable problem (Nagel, 1974, p. 439). We can never formulate a conceptual schema enabling us to intuit the subjective character of another's experience, which constitutes the *hard problem of consciousness*.

2.2 The Hard and Easy Problems of Consciousness

Chalmers (1995) once remarked that there is no singular problem of consciousness. There are many, but we can generally consider the unexplained phenomena of consciousness to fall within two categories. Namely, there are both the "easy" and "hard" problems of consciousness (Chalmers, 1995, pp. 1-2). The easy problems are those considered accessible to the standard methodologies of cognitive science (Chalmers, 1995, p. 2). Chalmers (1995) compiles a non-exhaustive list of examples that can be considered easy problems. These include attention, deliberate control of behavior, the ability to categorize and respond to environmental stimuli, the difference between wakefulness and sleep, and the reportability of mental states.

Conversely, the hard problems consistently resist such methods and seem particularly immune to attempts at philosophical and scientific reduction. Regarding p-consciousness, the really hard problem, according to Chalmers (1995), is the problem of experience itself and how it emerges (p.3). In other words, how is it that the physical composition of our brains gives rise to experience? Nothing about the brain and its neural substrates readily provides a clear answer. This issue is precisely what most philosophers of consciousness call the *hard problem*. To illustrate the hard problem of consciousness scientifically, we can turn to well-documented cases of blindsight. Patients with blindsight have extensive damage to their primary visual cortex. Their eyes and optic nerves remain perfectly intact, but the pathway through which visual information primarily travels within the mammalian brain is severely damaged. This damage is why they have no conscious recollection of being shown a visual stimulus. However, when asked to guess whether a visual stimulus was presented, they often get it right (Matlin, 2005, pp. 93-94). While the primary visual pathway is damaged, one possibility is that a secondary and evolutionarily much older one remains and processes the information pertaining to the stimulus.[3] However, since there is no recollection of having seen an image, we can conclude that the secondary pathway is not subject to conscious awareness and thus lacks any sort of phenomenal characteristics.[4] We can then ask, what about the physical composition or organization of the primary visual pathway in humans creates this qualitative difference? At this point, we may be tempted to say that the qualitative difference may reside in how the information is processed. However, there is nothing about information processing that necessarily entails subjective experience. After all, modern computers engage in some level of information processing, but we would not readily consider them p-conscious. Thus, we can see why the really hard problem persists.

How, then, may we go about explaining p-consciousness? One approach is to develop a feasible conceptual or mechanistic framework for understanding conscious phenomena. As mentioned earlier, such an approach would address the easy problem of consciousness (Chalmers, 1995, p. 2). Representational theories of consciousness attempt to generally explain p-consciousness by appealing to the representational contents of experience. Traditionally, the representational contents of p-consciousness have been divided into two categories: intentional

[3] Knudsen (2020) provides an excellent contrast between the two visual pathways used by animals. Mammals like us primarily rely on the retinogeniculate pathway (RGN) for processing visual information. Conversely, animals such as reptile, amphibian, and avian species use the retinotectal pathway which spans the optic tectum (OT). The mammalian equivalent of the OT is called the superior colliculus (SC). Given that we also possess this pathway, though no longer rely on it extensively, we can conclude it originated with our last common ancestor shared by these species (hence its relatively old age compared to the RGN).

[4] The debate for how blindsight functions neurologically is ongoing. Nonetheless, cases of blindsight remain relevant throughout the consciousness literature and have for some time (see Block, 1995). For some recently suggested readings concerning the role of alternative visual pathways in unconscious visual information processing, see Derrien et al. (2022) and Guo et al. (2024).

states (which include beliefs and desires) and non-mental perceptual states like emotions, pains, and other bodily sensations (Seager, 2016, p. 216). However, representationalism argues that the latter category can be subsumed under the first one (Seager, 2016, p. 216). Representational theories of consciousness come in many forms. However, for this thesis, I will only be introducing two broad categories, which are first-order and higher-order representational theories.[5] *First-order representational* theories (FOR) of consciousness try to account for all the felt properties of p-consciousness by cashing them out as representational contents stemming from initial percepts (Carruthers, 2000, pp. 257-258). According to FOR theories, we can distinguish the experiences of seeing red and green colors by deferring to differences in the properties represented within the mind, like the reflective properties of surfaces (Carruthers, 2000 pp. 257-258).

Furthermore, p-conscious experiences are poised to impact one's beliefs and reasoning processes to guide behaviors (Carruthers, 2000, p. 257). However, these theories do not move past first-order perceptual states to explain p-consciousness. Carruthers contends that FOR theories can easily provide an evolutionary explanation regarding p-consciousness (p. 258). He suggests that FOR theories just need to explain, using evolutionary terms, how the following transitions occurred sequentially (a-d): (a) organisms with basic reflexive locomotory behaviors, (b) complex behavioral schemes guided by incoming sensory information, and (c) organisms with a capacity for learning action-schemas; and then on to (d) organisms where perceptual information and contribute to the formation of concepts and reasoning (p. 4). Carruthers (2000) considers it obvious that evolutionary gains incurred on organisms come from an increasing array of flexible *behavioral repertoires* (p. 258). Moreover, according to FOR theories, any

[5] For an excellent and extensive overview of competing theories of consciousness, see Seager (2016).

evolutionary narrative concerning p-consciousness amounts to the evolution of perceptual experience (Carruthers, 2000, p. 258).

Conversely, *higher-order representational* accounts (HOR) attempt to explain p-consciousness in several ways, giving rise to competing accounts. However, they share some basic assumptions. According to HOR, p-consciousness is most accurately described as a variety of intentional or representational content, which play a role within the causal architecture of the mind (Carruthers & Gennaro, 2023, sec. 2). HOR theories further stipulate that in order for a representation to be p-conscious, it needs to be subjected to a higher representation of it. One variety of HOR are *higher-order experience theories* (HOEs) offered by proponents like Armstrong (1968) and Lycan (1996). HOE theories posit a set of inner scanners, which are directed onto our first-order mental states and cause them to become p-conscious (Carruthers, 2000, p. 260). The evolutionary question for these theories is how such an inner sense evolved. Carruthers (2000) states that these theories generally appeal to complex levels of organization within an organism, which play a causal role (p. 260). *Higher-order thought* theories (HOTs) operate quite differently. Instead, HOTs argue that p-consciousness is created whenever a first-order perceptual state is targeted by a higher-order thought (Carruthers, 2000, p. 258). According to Carruthers (2000), these theories come in two varieties: *actualist* and *dispositionalist.* For actualist accounts, the presence of a HOT is sufficient for p-consciousness; however, dispositionalist accounts state that the *availability* of perceptual states to a HOT is what matters (see Carruthers, 1996). A persistent complication with HOR theories of consciousness when providing an evolutionary account of p-consciousness is that they need to specify what high-order cognitive capacities need to evolve.

Feinberg and Mallatt (2020) posit a theory they call *neurobiological naturalism*, which is, at its core, a representational theory of consciousness. More specifically, their theory largely falls within the first-order category of representational theories and mostly considers first-order percepts of experience as requisites for p-consciousness. For example, Feinberg and Mallatt (2016) emphasize the role of visual perception and the first image-forming eyes in early vertebrates in the evolution of p-consciousness within encephalized descendant animal lineages. Feinberg and Mallatt traverse many of the details mentioned above by Carruthers (2000) when describing how first-order representational theories could fruitfully establish an evolutionary narrative of consciousness. For instance, Feinberg and Mallatt (2016) describe how comparatively primitive forms of microscopic life were likely limited to basic locomotory and reflexive behaviors. Under their view, the evolution of certain neurological components of sensory perception, like early nervous systems, meant that incoming sensory signals could modify behaviors to greater degrees of complexity. Then, learning and more complex forms of cognition culminated in the evolutionary development of the first brains and centralized nervous systems. Their account may nonetheless posit aspects of high-order theories when it comes to particularly complex forms of cognition like self-recognition and what they refer to as "meta-awareness," which they define as an awareness that one is aware (Feinberg & Mallatt, 2016, p. 215). However, in their account, they traverse an additional layer of complexity that many representational theories of consciousness do not address. The aforementioned representational theories of consciousness generally stay within the cognitive domain, whereas Feinberg and Mallatt do not. They attempt to establish an evolutionary narrative that details the neurobiological components they consider collectively sufficient for p-consciousness to have evolved in animals. In doing so, their theory becomes subjected to analytical difficulties that other purely cognitive theories of consciousness only sometimes encounter. For instance, if a theory specifies certain neurobiological components involved with p-consciousness, it must provide substantive reasons for thinking so.

Moreover, the theory should be able to distinguish p-conscious neurological components from those that are not p-conscious and explain what creates this difference. Otherwise, we risk arbitrarily assigning p-consciousness to certain neurological components. I will now summarize their theory before subjecting it to philosophical scrutiny.

3 FEINBERG AND MALLATT'S NEUOBIOLOGICAL NATURALISM

In this section, I will summarize Feinberg and Mallatt's (2020) theory of *neurobiological naturalism, which* encompasses what they consider to be the evolution of p-consciousness in animals. They aim to provide an account of the evolution of critical neurobiological components that they consider sufficient for p-consciousness to emerge. After summarizing their account, I will move on to a philosophical critique of their views in the next section.

3.1 Summary of Neurobiological Naturalism

In principle, many of life's basic processes remain explainable through modern biological, chemical, and physical sciences. However, when it comes to p-consciousness and its phenomenal properties, we seem to be at a loss for fully accommodating it into our scientific worldview. As mentioned before, there exists in our understanding of consciousness what philosopher Levine (1983) called an "explanatory gap" between its physical substrate (i.e., the brain) and its subjective aspects, the fact that it feels like anything at all. Feinberg and Mallatt (2020) set out to fill the gap with their *neurobiological naturalistic* approach, which incorporates a neuroevolutionary perspective and weak emergence theory. In doing so, they aim to demonstrate that p-consciousness can be explained in a way that does not require strong emergence (Feinberg & Mallatt, 2020, p.1). They track the evolutionary development of what they consider to be the *special neurobiological features* that are collectively sufficient for p-consciousness to emerge. These include, among others, elaborate sensory organs, neural hierarchies, reciprocal communication networks between neural processing centers, and processing centers for imagistic and affective consciousness (Feinberg & Mallatt, 2020, p. 6). An in-depth review of these concepts will follow towards the end of this section.

Emergence is a property of *complex systems* where novel properties, termed *system properties*, can result from the interactions of constituent parts, which are called *aggregate functions*. The phrase *aggregate functions* is used by proponents of emergence theory, like Bedau (1997), in an almost mathematical sense where system properties are a function (output) of the various micro-level interactions of the system's parts (input). Fluid dynamics, for instance, are an emergent system property of the aggregate interactions of water molecules (Bedau, 2002, p. 9). Generally, two distinguishable kinds of emergence theory are *weak* and *strong* emergence, and they make different fundamental assumptions. Where the former assumes ontological reducibility, the latter does not fully embrace it. Ontological reducibility, within emergence theory, is the notion that emergent system properties cannot be adequately explained by their parts in isolation but rather only by accounting for the properties of the parts and their interactions. In that sense, accepting ontological reducibility also buys into the notion of reductive physicalism, but with added steps to accommodate the complexity of dynamic systems. To use the earlier example of fluid dynamics, a single water molecule and its chemical properties are insufficient for the emergence of fluid dynamics, let alone to fully explain it. The interactions of large aggregates of water molecules cause fluid dynamics to emerge as a system property, and understanding their various molecular interactions lets us explain how fluid dynamics operate.

The extent to which fluid dynamics can be understood this way makes it ontologically reducible. This summary is, admittedly, a broad characterization of emergence theory, to which I can do no justice here. However, for this thesis, it will suffice to say that Feinberg and Mallatt's view of emergence and how it is incorporated within their work falls within the purview of *weak emergence*. Feinberg and Mallatt (2020) consider the special neurobiological features to have weakly emerged throughout evolution. Together, they are collectively sufficient for the subsequent emergence of p-consciousness within certain animal lineages. The philosophical treatments of emergence in the work of Bedau and Humphrey (2008) have largely influenced their work, which I will return to in the argumentative section (sec. 4). Now, we can discuss how Feinberg and Mallatt incorporate emergence into their view of p-consciousness and how it evolved in animals.

Feinberg and Mallatt treat p-consciousness as a weakly emergent system property of the brain and nervous system. Similar treatments exist in their previous works (Feinberg, 2012; Feinberg and Mallatt, 2016, 2018). Other philosophers of consciousness have also incorporated forms of emergence into their theories of consciousness. Chalmers (2006), for instance, considers p-consciousness to be an emergent property, albeit an irreducible one that entails a form of strong emergence. Since Feinberg and Mallatt consider p-consciousness to be weakly emergent, they assume that it will eventually be rendered fully understandable after all of its micro-level interactions within the brain and their causal relationships are understood. In other words, understanding all the features of the brain that create p-consciousness will fill in the explanatory gap. Feinberg and Mallatt reject the notion of any variety of strongly emergent features of p-consciousness.

3.2 Feinberg and Mallatt's Model of Emergence

Feinberg and Mallatt (2020) provide a model for the emergence of consciousness in animals, which consists of three levels. The first level, which they call *life*, details the evolution of life on our planet, from its simplest beginnings ~3.7 billion years ago[6], to profound levels of complexity. Some emergent structures of early life within the Earth's ancient oceans included simple organic macromolecules from which the first boundary membranes are thought to have

[6] Ancient seafloor deposits found in Quebec show signs of iron-oxidizing bacteria, dating to around 3.77 billion years ago (Dodd et al., 2017).

originated, which eventually led to the first embodied protocells (Feinberg & Mallatt, 2020, p. 6). According to England (2012), these protocells would form the basis for early microscopic life forms capable of metabolic processes and genetic reproduction through various means. According to Feinberg and Mallatt (2020), evolution proceeded over billions of years from the simplest single-celled organisms to the first multicellular Eukaryotic marine organisms (p. 6). Around 700-600 million years ago, the first marine animals evolved with fully specialized cells, organs, and nerve cells, which enabled basic locomotory behaviors (Feinberg & Mallatt, 2020, p.7).[7]

The second level of their model details the evolution of early nervous systems, reflexive behaviors, and brains. Based on comparisons of modern cnidarians (e.g., jellyfish) and other simple marine worm species, they deduce that early forms of complex animal life had rudimentary nervous systems (Feinberg & Mallatt, 2020, p. 7). These animals likely had distributed nerve networks (nerve nets) without anything resembling a centralized brain (Feinberg & Mallatt, 2020, p. 7). They would have possessed mechanoreceptors for touch sensations, chemoreceptors for discriminating tastes and smells, and photoreceptors sensitive to light (Feinberg & Mallatt, 2020, p. 7). However, ~580-520 million years ago, during the Cambrian period, many marine vertebrate and invertebrate species that were more similar to what we would recognize today as animals evolved.[8] In many of these early animal lineages, parts of their nerve nets increased in density and became enlarged for sensory information processing (Feinberg & Mallatt, 2020, p. 7). According to Feinberg and Mallatt (2020), these developments marked the beginning of the first "core brains" in animals that would evolve into a

[7] One example of a simple locomotory behavior is *phototaxis* where microscopic photosynthetic organisms can move towards sunlight using small hair-like cilia structures on their cell bodies.
[8] See Feinberg and Mallatt (2016, 2018) for their more detailed account of these evolutionary events.

centralized nervous system with a brain and spinal cord (p. 7). They consider more elaborate neural connections and animal behaviors to have emerged during this phase of evolutionary development and have made such remarks elsewhere (see Feinberg & Mallatt, 2016, 2018).

Their model's third and last level describes the emergence of the *special neurobiological features* of complex brains, which are collectively sufficient for p-consciousness. Feinberg and Mallatt consider these features to be their version of what has been called the neural correlates of consciousness (NCCs) by other researchers (see Koch, 2019; Seth & Bayne, 2022; Tononi et al., 2016). According to Feinberg and Mallatt (2020), most researchers derive their NCCs from an intensive study of the mammalian or human cerebral cortex (p. 7). Conversely, Feinberg and Mallatt (2020) derive their *special features from* two fundamental assumptions of theirs:

(1) "If an animal has neural pathways that carry mapped, point-by-point signals from the sensed environment, from different senses (e.g., vision, touch, hearing), and if these sensory maps converge in the brain, then that animal consciously experiences a unified, mapped, multisensory image of the environment" (p. 7).

(2) "If an animal shows complex operant learning, i.e., learning and remembering from *experience* to avoid harmful stimuli and to approach helpful stimuli, then that animal has negative and positive feelings of affective consciousness" (p.7).

The first assumption they make related to sensory mapping shows signs of first-order representationalism imbedded within their account because it deals with initial sensory percepts coming from the environment. Part of what it means to have experiences, for Feinberg and Mallatt, is synonymous with having sensory signals from the environment come together within the brain to form images of the environment. Furthermore, these images can be multisensory because animals often have more than one sensory modality like vision, hearing, taste, smell, and touch. Regarding their second assumption, Feinberg and Mallatt appear to assume that operant learning requires experience, and being able to avoid harmful stimuli (like toxins) and seek out positive ones (like food) requires affective mental states. From these two assumptions, they derive the following *special features*: (a) neural complexity, (b) elaborate sensory organs, (c) neural hierarchies with reciprocal communication networks, (d) pathways that create sensory maps of the environment, (e) mechanisms for selective attention and arousal, (f) as well as those for short and long term memory. How they derive these special features from their two assumptions related to sensory mapping and affective consciousness is methodological. Feinberg and Mallatt (2016) often begin by proposing what they consider to be a feasible set of conditions to meet for something like p-consciousness or affective states. Then, they ask what would be required neurobiologically for the conditions to be satisfied based on what is known about the brain and nervous system.

There are six features of emergence theory that Feinberg and Mallatt have incorporated into their account of how these special neurobiological features evolved in animals and caused p-consciousness to emerge, two of which have already been mentioned, which are *complex systems* and *aggregate functions*. Both can be related to *neural complexity*, special feature (a), where an increasingly complex arrangement of neural connectedness constitutes a complex system of various aggregate functions (e.g., the interactions of neuronal populations and the functions they collectively perform). The third feature is *a hierarchical arrangement* of different levels within a system (Feinberg & Mallatt, 2020, p. 3). For instance, if we view animals as biological systems, the cell can be considered the smallest relevant unit. One step up, large congregations of cells create tissues, which comprise organs, and then organs can together compose organ systems. Each step up represents a marked increase in complexity, and each step can be treated as a level within a biological hierarchy. Feinberg and Mallatt (2016) have called this *self-organization*, where the *self* refers to the embodied organism and *organization* relates to its biological constitution. Under their account, neural hierarchies (special feature (c)) are hierarchical arrangements of many levels within the nervous system. These have become increasingly complex through differentiating cell types and their various functions. The fourth feature of emergence they use is called *reciprocal connections*, which refers to circular causality (Feinberg & Mallatt, 2020, p. 3). Circular causality occurs when the lower levels of a hierarchy influence change at higher levels, creating changes at the lower levels. Homeostasis, the process by which organisms regulate the internal chemistry of their bodies to fall within an optimal range (e.g., body temperature), represents a circular causality necessary for complex animal life. The degree of interconnectivity required to facilitate these reciprocal communication networks for homeostasis throughout nervous systems also constitutes neural hierarchies, adding to their complexity. *Constraint* is the next feature, where organization at the highest levels limits what its constituent parts at lower levels can do (Feinberg & Mallatt, 2020, p. 3). For example, the human body cannot survive if some of its cells begin depriving others of vital resources, like in the case of cancer (Feinberg & Mallatt, 2020, p. 3). Lastly, *multiple routes* are another feature of emergence they use, akin to the concept of multiple realizability regarding how a single mind state can be realized in multiple ways (see Bickle, 2019). Bedau (2008) called this feature "macro explanatory autonomy," where an emergent system property can be realized from different sets of its constituent parts and their interactions (pp. 181-182).

Feinberg and Mallatt (2020) do not provide an in-depth overview of their *special features,* which are said to have weakly emerged from and are the consequence of evolutionary development within certain animal lineages. However, they have elsewhere, which I will now attempt to summarize. Under their account, (a) *neural complexity* entails a process where nerve cells become increasingly numerous over time and differentiate into many cell types, all capable of sensory information processing and sensitive to different stimuli. Neural complexity is also involved in the formation of neural hierarchies. (b) *Sensory organs*, such as the eyes, are necessary for receiving and relaying incoming sensory information to the brain. According to Feinberg and Mallatt (2016, 2020), (c) *neural hierarchies* display both nested and non-nested characteristics. They are nested in the sense that their lower and higher levels are composed of the same basic components (e.g., neurons) but are also non-nested because there is a degree of physical separation between some levels as they are spread throughout the body (Feinberg & Mallatt, 2016, pp. 28–29). According to them, a major function of neural hierarchies is non-nested, which they call *topical convergence* (Feinberg & Mallatt, 2016, p. 29). This function involves the upward flow of sensory information from low-level neurons to higher levels. To illustrate this point, Feinberg and Mallatt (2016) use the following example:

"In our visual processing…low-level neurons have less differentiated response characteristics (responding, for example, to viewed points or short lines) and they project to higher neurons that have increasingly specific responses. Finally, in the cerebrum's temporal lobe…certain neurons respond to the most highly integrated visual stimuli" (p. 29).

These highly specialized neuronal cells are often referred to as grandmother cells because some of them may quite literally only respond to the visual stimuli of your grandmother's face (Feinberg & Mallatt, 2016, p. 29).[9] (d) *Sensory pathways* within the brain engage in a process Feinberg and Mallatt (2016) call *topographic representation*. This concept refers to the spatial distribution on receptor cell surfaces, like the retina, having a corresponding sensory region

[9] I have included this example only to illustrate further what they call topical convergence. However, the grandmother cell characterization of specialized neurons is not uncontested within neuroscience (see Bowers 2011).

within the brain (Feinberg & Mallatt, 2016, p. 24). These representations are often referred to as maps because the connectivity between receptor cell surfaces and the brain's sensory regions results in sensory inputs mapping onto those corresponding regions. In a way, this process can also be described as a mental mapping of external environments. Many researchers consider maps ontologically necessary aspects of animal consciousness (Edelman, 1992; Damasio, 1999; Feinberg, 2012). (e) *Selective attention mechanisms* in the brain have been defined by Feinberg and Mallatt (2018) as being utilized by animals for "focusing consciousness onto salient objects in the environment," where arousal operates as a related mechanism for adjusting object salience (p.69). (f) Lastly, they claim that *short-term memory* is minimally required for the continuity of experience. In contrast, long-term *memory* represents a higher capacity which likely evolved shortly after p-consciousness evolved (Feinberg & Mallatt, 2018, p. 69).

Under their view, the *special neurobiological features* of consciousness (a-f) emerged naturally from evolutionary processes. They date the emergence of the *special features* as far back as the "Cambrian explosion" 540-500 million years ago. They hypothesize that the special features that are collectively sufficient for p-consciousness evolved because they lent competitive advantages to organisms that had it related to enhanced sensorimotor capabilities. Feinberg and Mallatt (2016) hypothesized that there were strong selective pressures stemming from the first predator-prey dynamics that shaped the development of elaborate sensory organs like the first image-forming eyes. These events increased behavioral complexity related to foraging, predation, and predatory aversion on behalf of prey species (Feinberg & Mallatt, 2016, pp. 62-63). The competitive intensity of early to late Cambrian marine ecosystems was likely extreme, and accurate sensory perception would have been incredibly advantageous for most marine animals alive at that time. However, these remarks are merely theoretical stipulations, and the debate over how these traits evolved and why is ongoing. For this thesis, I will focus on their argument for how these features can be used to fill in the explanatory gap in the next section.

4 PHILOSOPHICAL CRITIQUE

In this section, I will summarize what I call Feinberg and Mallatt's *special features argument,* which they claim dissolves the hard problem of p-consciousness (Feinberg & Mallatt, 2020, pp. 9-12; see also Feinberg, 2011, pp. 30-32). I will argue that their argument rests on fundamental assumptions that invariably confuse the hard and easy problems of consciousness. Finally, I consider two ways out of this mistake of theirs which would be internally consistent with the rest of their theory and how they hold up to philosophical scrutiny.

4.1 Special Features Argument

Feinberg and Mallatt (2020) contend that while they do not consider there to be any "scientific explanatory gap" between the brain and how p-consciousness emerges from the perspective of neurobiological science, there remains an "experiential gap" (p. 12). They describe the *experiential gap* as a gap in our understanding between scientific explanation and the subjective experience of consciousness (Feinberg & Mallatt, 2020, p. 12). However, this definition corresponds to what Levine (1983) originally called the explanatory gap between our understanding of the brain and how it creates subjective experience. As mentioned earlier, the hard problem of consciousness, as specified by Chalmers (1995) is the problem of how subjective experience (i.e., p-consciousness) is created by the physical composition of our brains. Feinberg and Mallatt further illustrate what they consider to be the experiential gap of consciousness by referencing C.D. Broad's (1925) thought experiment of an omniscient mathematical archangel. He argued that even if such a being were fully knowledgeable of the chemistry of ammonia, they would nonetheless fail to predict how it would smell. C.D. Broad's thought experiment of a mathematical archangel is a rejection of reductive physicalism, which operates in much the same fashion as Frank Jacksons' (1986) knowledge argument.

Reductive physicalism is the notion that everything we experience is ultimately physical and can be reduced to its physical basis. There are competing varieties of reductive physicalism that make different assumptions (see Stoljar, 2024). However, within the purview of Feinberg and Mallatt's work and this thesis, this definition will suffice. These lines of argumentation, offered by C.D. Broad and Jackson, aim to demonstrate that physicalism does not fully explain the irreducibility of phenomenal experiences. We're left to conclude that Feinberg and Mallatt's use of the phrase experiential gap in place of the explanatory gap (i.e., the hard problem), while awkward, means essentially the same thing.

Feinberg and Mallatt (2020) attempt to resolve the experiential gap without necessitating a dualistic understanding of mind and body to be compatible with ontological reductionism and their naturalistic outlook (p.12). Their concern is that without a neurobiological level of analysis describing the physical basis of p-consciousness, a dualistic treatment of p-consciousness would still be possible. The mind and body could be treated as separate things, where one is not reducible to the other. Moreover, they want to avoid invoking any strongly emergent features to explain p-consciousness. As mentioned earlier, to say a phenomenon is strongly emergent is to say that it can never, not even in principle, be reduced to its constituent micro-level interactions. Strong emergence is incompatible with the notion of ontological reducibility embraced by weak emergence theory. I call their argument the *special features argument*, which can be summarized as follows:

"If it is true, as we propose, that the *personal life* of any embodied organism is an emergent process of a physical system…then subjectivity is a critical but *biologically natural* element of what we experience as a phenomenal state; and if it is true, as we propose, that the addition of the special neurobiological features of complex brains…provides the *biologically* *natural* elements necessary for the hierarchical emergence of phenomenal consciousness, then we have enumerated all the prerequisites that are required for the natural emergence of subjective experience" (Feinberg & Mallatt, 2020, p.12).

Their special features argument operates as a series of conditionals, which can be disentangled and engaged with individually. The first conditional is what I will call *embodied subjectivity* (ES):

"If it is true, as we propose, that the *personal life* of any embodied organism is an emergent process of a physical system…then subjectivity is a critical but *biologically natural* element of what we experience as a phenomenal state" (Feinberg & Mallatt, 2020, p. 12).

This conditional considers embodiment a weakly emergent process of a physical system, which Feinberg and Mallatt (2020) consider sufficient (provided the other special features are present) for organisms to have subjective experiences (p. 6). The next conditional can be called *hierarchical emergence of p-consciousness* (HEP):

"[If] it is true, as we propose, that the addition of the special neurobiological features of complex brains…provides the *biologically natural* elements necessary for the hierarchical emergence of phenomenal consciousness, then we have enumerated all the prerequisites that are required for the natural emergence of subjective experience" (Feinberg & Mallatt, 2020, p.12).

Provided the special neurobiological features of a complex brain are present within an embodied organism. Feinberg and Mallatt consider all the conditions for the emergence of p-consciousness to be met. All of their criteria (embodiment and special features) are considered to be *natural* in the sense that they weakly emerged from a physical system as system properties. With this account, they establish a clear directionality for the evolution of p-consciousness in animals: *Life + Special neurobiological features* → *p-consciousness* (Feinberg & Mallatt, 2020, p.12).

Feinberg and Mallatt (2020) claim that with their naturalized account of the physical emergence of p-consciousness, the distinction between "being and experiencing versus observing and describing" (i.e., experiential gap) is accounted for by p-consciousness (p. 12). To clarify what they mean, consider a thought experiment proposed by Globus (1973) involving a fictional "autocerebroscope," which allows you to observe neural activity within your brain related to any particular experience. Their phrase "being and experiencing" is meant to encapsulate the subjectivity of that experience, whereas "observing and describing" captures the seeming objectivity of the observations you make. Insofar as we may struggle to explain why such neural activity feels like anything, Feinberg and Mallatt (2020) claim to have resolved this discrepancy with their account. Furthermore, they claim that p-consciousness "…has a scientific explanation that adheres to and is consistent with the principles of emergence in the rest of nature" (p.12). Insofar as their explanation of p-consciousness fits all the criteria they consider consistent with the principles of weak emergence, they claim that p-consciousness does not entail any sort of strong emergence (Feinberg & Mallatt, 2020, p. 12).

4.2 Objections and Counter Objections

On its face, this line of argumentation is problematic for a few reasons. First, the special features argument operates as a series of conditionals which presuppose a set of criteria that occur naturally through emergence, like the embodiment of organisms, which together entail p-consciousness (ES).[10] Second, these presuppositions are then used as justifications for having

[10] To review, the special features are (a) neural complexity, (b) elaborate sensory organs, (c) neural hierarchies with reciprocal communication networks, (d) pathways that create sensory maps of the environment, (e) mechanisms for selective attention and arousal, (f) as well as those for short and long term memory.

resolved the hard problem posed by p-consciousness by positing the emergence of special neurobiological features (HEP). Third, an additional assumption they make is that providing a naturalistic account of p-consciousness evaporates the mystery of its operations. Instead, it seems they have used p-consciousness itself to explain away the explanatory gap it creates (i.e., their experiential gap) which presents a vicious circularity. How did this mistake of theirs occur in the first place? I would like to suggest that Feinberg and Mallatt have confused the hard and easy problems of consciousness.

Chalmers (1995) considers the easy problems to include aspects of attention, deliberate control of behavior, the ability to categorize and respond to environmental stimuli, the difference between wakefulness and sleep, and the reportability of mental states, among others (p. 2). These aspects of cognition are generally accessible to the standard methodologies of cognitive science. He considers it a conceptual fact, which is true by necessity or definition, that these cognitive phenomena and their explanations only require a level of explanation specifying their various functions (Chalmers, 1995, p. 4). In other words, they are functionally definable, and as such, they can be defined and explained functionally (Chalmers, 1995, p.4). Problems within cognitive science, like learning, can be approached by studying how incoming environmental stimuli modify behaviors. If we can further show how these changes relate to an underlying neural or computational mechanism, we have successfully explained learning (Chalmers, 1995, p. 5). We can carry out this investigative operation on many cognitive phenomena that are perhaps within the vicinity of p-consciousness to explain their performance and functional characteristics.[11] However, once we approach the question of why performance and functional characteristics coincide with experience, our ordinary investigative practices fail

[11] Chalmers (1995) provides a list of examples such as "perceptual discrimination, categorization ability, internal access, and verbal report" (p. 5).

(Chalmers, 1995, p. 5). According to Chalmers, simple, functional explanations still leave the question of experience completely open (p. 5).

Feinberg and Mallatt's neurobiological naturalism primarily operates at a functional level of description, albeit with an account of how these functional characteristics evolved within animal lineages. Their account operates at a functional level insofar as their approach explains what neurobiological features are required for p-consciousness by defining their functional roles in the emergence of experience. It seems, then, that their account falls within the purview of the easy problems and does not approach the hard problem, much less cause it to dissolve. Any consideration of theirs for having done otherwise represents a conflation of terms.

There are two ways Feinberg and Mallatt could respond to these criticisms and remain internally consistent with the rest of their theory. First, they could simply abandon the notion that their naturalistic account solves the hard problem. Instead, they could claim that their account represents a solution to some of the easy problems, like those that pertain to learning, attentional states, and responding to environmental stimuli, to cite some examples. Second, they could sign onto the notion of an epistemic variety of *virtual irreducibility*, which is consistent with the literature on weak emergence and the principle of ontological reducibility. My use of the phrase *epistemic variety* is meant to refer to a classification of irreducibility, which stems from our limited capacity to have complete knowledge of a dynamic system due to the formidable computational complexity of understanding future iterations of it. The term *virtual* is meant to be similar to the claim that something is very nearly the case. In that sense, virtual irreducibility is synonymous with Bedau's (1997) concept of incompressibility.

In principle, and according to weak emergence theory and Bedau's incompressibility, we can know everything about a system property at time *T* if we can account for all of its micro-level interactions. However, doing so never tells you anything about future states of that system property, but this isn't to say that it would be impossible, just computationally very complex (see Bedau, 1997, 2008). As a result, while something like p-consciousness may not be irreducible in the sense that it cannot be reduced to the micro-level interactions of all its neurobiological components, doing so does not give us full knowledge over all future iterations of it as an emergent system property. Such knowledge would require formidably high levels of computational complexity to obtain and be nearly unapproachable as a result. In principle, however, the computational task could be undertaken. Hence the phrase 'epistemic variety of virtual irreducibility' is meant to capture the appearance of irreducibility presented by enormous computational complexity. They could claim that their account can explain the irreducibility of p-consciousness by appealing to this notion of an epistemic variety of virtual irreducibility. To say something is virtually irreducible is to make an epistemic claim regarding our knowledge of an emergent system property and its limitations in the case of p-consciousness. An appeal to our epistemic limitations over a system property does not necessarily contradict ontological reducibility because it remains, in principle, feasible although computationally very complex. For Feinberg and Mallatt, the hope is that by appealing to this notion of virtual irreducibility as an epistemic claim, the rest of their naturalistic account of p-consciousness as a weakly emergent system property would remain intact because it would not invoke features of strong emergence. To say that an emergent system property, like p-consciousness, is strongly emergent is to say that it could never, not even in principle, be reduced to its constituent micro-level interactions.

The first option seems less problematic than the second. It amounts to nothing more than Feinberg and Mallatt falling back on their ample functional descriptions of what they consider critical components sufficient for animal p-consciousness. They can do this without claiming to have solved the hard problem outright because their level of analysis simply approaches the easy problems. These descriptions are accompanied by an evolutionary narrative they establish concerning the developments of functional components, like elaborate sensory organs. The problem may exist within their proposed narrative. There are, for instance, competing narratives for how these features evolved and why (see Carruthers, 2000; Budd, 2015). However, engaging with these alternative views is beyond the purview of this thesis. Feinberg and Mallatt may, however, be unsatisfied by this approach. Perhaps they feel that once you have provided a satisfactory functional account of p-consciousness, then nothing more is required to explain it. Such sentiments would be understandable; however, we must not confuse descriptions of functional mechanisms with an explanation of why those mechanisms necessarily feel like anything. In other words, describing how a cognitive mechanism, like attention, works differs from explaining why it necessarily entails phenomenal experience. Moreover, we cannot forget that with any functional evolutionary account, we may invariably assume that the trait in question evolved to have an adaptive function but overlook the possibility that it may simply be a byproduct of another evolved trait with a different function. In other words, something like p-consciousness could be nothing more than what Gould and Lewontin (1979) have called spandrels.

The second option can be problematic because it seems to appeal to something like mysterianism, which is the view that the hard problem is simply too hard for us to solve with our limited cognitive capacities (Kriegel, 2007, p. 36). McGinn (1989) referred to the hard problem of p-consciousness as “cognitively closed” to us for the same reason. Chalmers (1995), however, considered these sentiments to be premature (p. 13).

The notion of an epistemic variety of virtual irreducibility stems from the seemingly overwhelming complexity of the computational task involved with deciphering all future states of p-consciousness as a dynamic, weakly emergent system property. Such a notion certainly appears, on its face, to be quite mysterious. Nonetheless, this maneuver may only succeed if it aims to avoid a direct appeal to strong emergent features of p-consciousness. However, that does not mean they have provided a more workable alternative hypothesis. Ideally, if one were to eliminate the possibility of accepting a null hypothesis (i.e., the hypothesis you are trying to disprove), you would be minimally required to provide a more workable alternative hypothesis. If a core tenet of virtual irreducibility is its epistemic unapproachability stemming from absurd levels of computational complexity, then the hypothesis would fail to be more workable than an appeal to strong emergence. Option two may be problematic, in that sense, because it provides very little reason to consider p-consciousness a weakly rather than strongly emergent property. In other words, option two may provide little reason to sign onto the notion of p-consciousness as ontologically reducible (weakly emergent) rather than irreducible (strongly emergent). To illustrate what I mean, consider the claim that p-consciousness is strongly emergent implies that it can never, not even in principle, be reduced to its constituent micro-level interactions, which is incompatible with the notion of ontological reducibility embraced by weak emergence theory. In that sense, p-consciousness, viewed as a strongly emergent system property that is irreducible, represents an *intelligibility problem*. Here, I use the phrase *intelligibility problem* to refer to the sort of epistemic limitations one encounters and is often confronted by in the presence of irreducibilities. We have problems understanding them to the extent that they resist reduction and simplification. However, suppose the reasons we are given for accepting that p-consciousness is ontologically reducible are more or less attached to this idea of the *appearance* of irreducibility, in the case of virtual irreducibility, because of certain epistemic limitations related to computational complexity. In that case, we are presented with little more than another kind of intelligibility problem and have swapped one for another with a different flavor. At that point, we are left with little recourse other than to defer to preference when accepting p-consciousness as a strongly or weakly emergent system property.

5 CONCLUSION

Throughout this thesis, I presented Feinberg and Mallatt's (2020) theory of neurobiological naturalism and subjected it to some philosophical criticisms. I argued that their claim to have filled the explanatory gap was based on a fundamental misunderstanding, which confused the hard and easy problems of consciousness. I presented two responses Feinberg and Mallatt could reasonably provide and demonstrated why they are both problematic. Nonetheless, I would like to make some further remarks. Namely, scientific accounts of p-consciousness like Feinberg and Mallatt's may be particularly susceptible to making this mistake of confusing the hard and easy problems. In order to avoid this mistake, we must be aware of an important distinction. Providing a functionally definable mechanism of how aspects of p-consciousness may operate is not, nor should it be, synonymous with having solved the hard problem. Furthermore, evolutionary narratives that attempt to describe how these functionally definable mechanisms evolved within animal lineages cannot overlook why these mechanisms should feel like anything at all, as opposed to occurring in the dark like other aspects of animal cognition.

REFERENCES

Armstrong, D. M. (1968). *A materialist theory of the mind*. https://ci.nii.ac.jp/ncid/BA21478325

Bayne, T., Seth, A. K., Massimini, M., Shepherd, J., Cleeremans, A., Fleming, S. M., Malach, R., Mattingley, J. B., Menon, D. K., Owen, A. M., Peters, M. A. K., Razi, A., & Mudrik, L. (2024). Tests for consciousness in humans and beyond. *Trends in Cognitive Sciences*, *28*(5), 454–466. https://doi.org/10.1016/j.tics.2024.01.010

Bedau, M. A. (1997). Weak emergence. *Noûs/NoûS*, *31*(s11), 375–399. https://doi.org/10.1111/0029-4624.31.s11.17

Bedau, M. A. (2002). Downward Causation and the Autonomy of Weak Emergence. *Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC)*, *6*(1), 5–50. https://dialnet.unirioja.es/descarga/articulo/5251315.pdf

Bedau, M. A. (2008). Is weak emergence just in the mind? *Minds and Machines*, *18*(4), 443–459. https://doi.org/10.1007/s11023-008-9122-6

Bedau, M. A., & Humphreys, P. (Eds.). (2008). *Emergence: Contemporary readings in philosophy and science*. MIT press.

Bickle, J. (2019). Multiple Realizability. Encyclopedia of Cognitive Science. Available online at: https://plato.stanford.edu/archives/spr2019/entries/multiplerealizability

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247. https://doi.org/10.1017/s0140525x00038188

Bowers, J. S. (2011). What is a grandmother cell? And how would you know if you found one? *Connection Science*, *23*(2), 91–95. https://doi.org/10.1080/09540091.2011.568608

Broad, C. (1925). *The Mind and its Place in Nature*. Routledge.

Budd, G. E. (2015). Early animal evolution and the origins of nervous systems. *Philosophical Transactions - Royal Society. Biological Sciences*, *370*(1684), 20150037. https://doi.org/10.1098/rstb.2015.0037

Carruthers, P. (1996). *Language, thought and consciousness*. https://doi.org/10.1017/cbo9780511583360

Carruthers, P. (2000). The evolution of consciousness. In *Evolution and the human mind: Modularity, language and meta-cognition*, ed. P. Carruthers and A. Chamberlain, 254-275. Cambridge: Cambridge University Press. https://philpapers.org/rec/CAREAT-20 and https://faculty.philosophy.umd.edu/pcarruthers/Evolution-of-consciousness.htm

Carruthers, P. (2003). *Phenomenal consciousness: A naturalistic theory*. Cambridge University Press (paperback edition).

Carruthers, P. & Gennaro, R. "Higher-Order Theories of Consciousness," *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), https://plato.stanford.edu/archives/fall2023/entries/consciousness-higher/

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 1–25.

Chalmers, D. J. (1998). *Facing up to the problem of consciousness*. Routledge (pp. 207–228). https://doi.org/10.4324/9780203826430-11

Chalmers, D. J. (2006). "Strong and weak emergence," in *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*, eds P. Clayton and P. Davies (Oxford: Oxford University Press), 244–254.

Damasio, A. R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Houghton Mifflin.

Derrien, D., Garric, C., Sergent, C., & Chokron, S. (2022). The nature of blindsight: implications for current theories of consciousness. *Neuroscience of Consciousness*, *2022*(1), https://doi.org/10.1093/nc/niab043

Dodd, M. S., Papineau, D., Grenne, T., Slack, J. F., Rittner, M., Pirajno, F., ... & Little, C. T. (2017). Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature*, *543*(7643), 60-64.

Edelman, G. M. (1992). *Bright air, brilliant fire*. New York, NY, USA: Basic Books.

England, J. L. (2013). Statistical physics of self-replication. *The Journal of Chemical Physics*, *139*(12). https://doi.org/10.1063/1.4818538

Feinberg, T. E. (2012). Neuroontology, neurobiological naturalism, and consciousness: A challenge to scientific reduction and a solution. *Physics of Life Reviews*, *9*(1), 13–34. https://doi.org/10.1016/j.plrev.2011.10.019

Feinberg, T. E., & Mallatt, J. (2020). Phenomenal consciousness and emergence: eliminating the explanatory gap. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01041

Feinberg, T. E., & Mallatt, J. M. (2016). *The ancient origins of consciousness*. MIT Press. https://doi.org/10.7551/mitpress/10714.001.0001

Feinberg, T. E., & Mallatt, J. M. (2018a). *Consciousness demystified*. MIT Press. https://doi.org/10.7551/mitpress/11793.001.0001

Feinberg, T. E., & Mallatt, J. M. (2018b). *Consciousness demystified*. MIT Press.

Globus, G. G. (1973). Unexpected symmetries in the "World Knot." *Science*, *180*(4091), 1129–1136. https://doi.org/10.1126/science.180.4091.1129

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Conceptual Issues in Evolutionary Biology*, *205*, 79.

Guo, F., Zou, J., Wang, Y., Fang, B., Zhou, H., Wang, D., ... & Zhang, P. (2024). Human subcortical pathways automatically detect collision trajectory without attention and awareness. *PLOS Biology*, *22*(1), e3002375.

Jackson, F. (1986). What Mary didn't know. *the Journal of Philosophy/the Journal of Philosophy*, *83*(5), 291. https://doi.org/10.2307/2026143

Koch, C. (2019). *The feeling of life itself: Why Consciousness Is Widespread but Can't Be Computed*. MIT Press.

Knudsen, E. I. (2020). Evolution of neural processing for visual perception in vertebrates. *Journal of Comparative Neurology*, *528*(17), 2888–2901. https://doi.org/10.1002/cne.24871

Kriegel, U. (2007). Philosophical Theories of Consciousness: Contemporary Western Perspectives. In *Cambridge Handbook of Consciousness* (pp. 35–36). Cambridge. https://doi.org/10.1017/cbo9780511816789

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

Lycan, W. G. (1996). *Consciousness and experience*. MIT Press.

Matlin, M. W. (2005). *Cognition*. John Wiley & Sons.

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

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, *83*(4), 435–450. https://doi.org/10.2307/2183914

Pitt, David, "Mental Representation," *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), https://plato.stanford.edu/archives/fall2022/entries/mental-representation/

Seager, W. (2016). *Theories of consciousness: An Introduction and Assessment*. Routledge.

Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, *23*(7), 439-452.

Squazzoni, F. (2008). *Emergence: Contemporary Readings in Philosophy and Science (Bradford Books) Bedau, Mark A and Humphreys, Paul (Eds.),* MIT Press: London, 2007 (Vol. 11, Issue 4, pp. 1–3). https://air.unimi.it/handle/2434/661758

Stoljar, Daniel, "Physicalism," The Stanford Encyclopedia of Philosophy (Spring 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), https://plato. stanford.edu/archives/spr2024/emtries/physicalism/

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nature Reviews. Neuroscience*, *17*(7), 450–461. https://doi.org/10.1038/nrn.2016.44