UNIVERSITY
OF OSLO

Master thesis

Physicalism and the Knowledge Argument:

An Argument for Higher-Dimensional Physicalism

Øystein Austaråa Rene

MA Philosophy

60 Credits

Department of Philosophy, Classics, History of Art and Ideas

Faculty of Humanities

Abstract

This paper explores a theory-based physicalism conditional on string theoretical frameworks in modern physics, with an aim to provide a novel response to Frank Jackson´s Knowledge Argument against physicalism. This presents an argument for what will be called Higher Dimensional Physicalism, which will help to explain the private nature of the phenomenal property of consciousness. The argument supposes that string theoretical properties, understood as physical facts, can provide a priori entailment of the phenomenal properties associated with qualia, which are interpreted as being instantiated in and constituted by the empirically unobservable spatial dimensions conjectured in modern physics. This is argued to represent a viable physicalist monism consistent with scientific ontology, to explain the private nature of phenomenal properties which themselves are empirically unobservable and knowable in practice only from the first-personal perspective.

Acknowledgments

First and foremost, I wish to express my sincere gratitude to my supervisor, Hedda Hassel Mørch. She has far exceeded all expectation with the generous devotion of her time and efforts in her careful analyses of my numerous rough drafts. Her thorough feedback has greatly enhanced the structure of my thesis through its development and helped to provide greater clarity of argument.

I thank my former lecturer and researcher, Andre Sevenius Nilsen, for taking the time to share relevant literature. His stimulating conversations on the topic in and outside of seminars has spurred my interest in the general field of Consciousness Studies as a whole.

I would also like to express my appreciation for my peers at the institute. Their inputs, encouragement, and good company in general has made the at times arduous process of dissertation writing seem far more sufferable.

Last but not least, I wish to thank my family for their unconditional support and encouragement.

Table of Contents

Introduction

The aim of this thesis is to defend a theory-based physicalism on consciousness, by providing a novel response to Jackson´s (1982) knowledge argument against physicalism. The basic concern will be addressing the private nature of phenomenal consciousness, where the private nature of phenomenal consciousness refers to how it is accessible only from the first-person perspective. I will do this by invoking a multidimensional spatiotemporal framework as that described by string theory or M-Theory, suggesting that the qualia Jackson discusses are constituted by string-theoretical facts in physics, thus considered as physical facts of Nature. The qualia Jackson discusses, then, will be argued as instantiated in and constituted by the higher spatial dimensions beyond those amenable to empirical observation, which can provide a theoretical physicalist account to provide a priori entailment of the phenomenal properties associated with qualia and how these fit in with a general scheme of Nature.

While Jackson argues qualia cannot be entailed by physical facts on the basis that qualia can only be known in practice from the first-person perspective, I argue that we can provide a priori entailment of qualia in principle, as with a third-person theoretical account, whereby qualia become understood as extensive in the higher physical dimensions posited by string theory/M-Theory.

This will be the point of contention with Jackson´s Knowledge Argument against physicalism, then. Based on his argument, the kind of epistemic gap Jackson argues exist between qualia and physical facts led him to suppose an ontological gap between qualia and physical facts *in principle,* which submits that we cannot deduce facts of qualia from physical facts even on a third-person account. In turn, this led him to suppose a dualist interpretation in the form of epiphenomenalism. This turn from an epistemic gap between third-personal accounts, i.e., physical descriptions of facts about consciousness, and the first-personal perspective of qualia *in practice*, which is then taken to imply an ontological gap between physical facts and qualia *in principle* is seen to equivocate on the relevant forms of knowledge. The claim that we cannot have knowledge in principle on phenomenal properties associated with qualia is what I will aim to refute.

Further, the knowledge of qualia in practice, I would qualify as a pre-theoretical, pre-conceptual form of knowledge, since having the sensational awareness of the phenomenal property of *redness,* for example, does not rely on any theoretical or conceptual framework to

classify or interpret the experience as such. This point will be addressed later in chapter 1 when looking at some previous objections to the Knowledge Argument raised by Lewis and Churchland.

In general, a theory-based physicalism can be standardly defined as saying, "a property is physical if and only if it is the sort of property that physical theory tells us about." (Stoljar, 2021). Thus, a theory-based physicalism applied to the phenomenon of consciousness, like the one I will argue for here, submits that a theory about physical facts (i.e., a third-person account) are exhaustively descriptive of facts about consciousness. More precisely, physical properties are those exhaustively described by some physical theory, such as fundamental physics.[1] Theory-based physicalism, in turn, says that all properties – or facts, states, events, processes, or similar – are either exhaustively described by physical theory or, to also capture macroscopic phenomena not directly described by physics, *constituted* by some such physical properties. In this case, the theory-based physicalism to be tried is conditional on string theory/M-Theory or some similar multidimensional theory bearing fruit in the domain of physics, as mentioned. The wider physical descriptions this provides, and how it relates to the target phenomenon ´consciousness´ and its phenomenal properties, will be developed further in chapter 3.

As previously defined, the private nature of phenomenal consciousness is accessible only from a first-person perspective, characteristic of the features of mind picked out by *qualia*. These phenomenal properties, and the mental features they refer to, will be argued for as an extension of the brain and its neural architecture. The way in which the brain and neural architecture give rise to consciousness states, to qualia, is describable on the neurobiological level of analysis in terms of the neural correlates of consciousness (NCC).

The NCC are standardly defined as "the minimum neuronal mechanisms jointly sufficient for any one specific conscious percept." (Koch, Massimini, et al., 2016). Further, a "full NCC", meaning that which is required to generate consciousness in general, is defined as "the neural substrates supporting conscious experiences in their entirety, irrespective of their specific contents." (Koch, Massimini, et al., 2016). Based on the accruing data of the relevant neuronal mechanisms, researchers can thus develop quantitative indices associated with conscious states. I will discuss this further in chapter 2 where I argue that standard descriptions of NCC are not sufficient to entail phenomenal consciousness since these only

[1] One might also add, e.g., chemistry, if one does not regard it as reducible to physics. Importantly, however, physical theory is not taken to include theories regarding those directly mentioning consciousness, such as psychology.

report on Access-Consciousness states (Block, 1995) which is a term I will define in chapter 1, as part of my response against Lewis´s objection to the Knowledge Argument.

The basic claim of my thesis argument is that phenomenal properties of consciousness are instantiated in and constituted by string theoretical properties in higher spatial dimensions. Further, the quantitative indices the NCC supply, based on neurobiological mechanisms in the brain, are to be regarded as only a cross-section of the overall structure of consciousness taken as a physical entity. This is because the brain and the NCC is the part which is amenable to empirical observation, obtaining in four dimensions of space and time, above those unobservable spatial dimensions string theory/M-Theory posit. This cross-section, the empirically observable mechanisms of the brain, do not entail qualia as such, argued here as instantiated in and constituted by physical properties in the higher spatial dimensions discussed above.

The standard interpretation of what the NCC are taken to include, see chapter 2, will be regarded here as *Four-Dimensional physicalism.* I call it Four-Dimensional Physicalism since studies of the NCC tend to only entail four-dimensional physics or classical physics, which does not posit any such higher physical dimensions *qua* string theory/M-Theory or, typically, quantum mechanics.[2]

Now, according to the Knowledge Argument, discussed in chapter 1, to the extent that Mary only has complete and perfect knowledge of four-dimensional physical descriptions, my claim is that the Knowledge Argument holds as a refutation of physicalism. This is so, since, as I will argue, four-dimensional physical information does not provide a priori entailment of qualia as mentioned above.

In chapter 1, I will begin by addressing the Knowledge Argument against physicalism, while looking at some previous objections mounted by Lewis and Churchland. If such responses fail to refute the knowledge argument in full, the details of their objections will help to show the motivation for my alternative response, leading to the broader theory-based physicalism to be argued here which include the additional spatial dimensions conjectured in string theory/M-Theory.

For short, we can refer to this as *Higher Dimensional Physicalism*. Importantly, this Higher Dimensional Physicalism entails Four-Dimensional Physicalism, since the physical

[2] The inclusion of quantum mechanical principles as part of the full NCC, considered as relevant neuronal mechanisms to support consciousness in general, is important for the plausibility of my thesis argument, due to the quantum mechanical principles governing the behaviour of strings in String Theory/M-Theory, as will be discussed in chapter 4.

facts posited/presupposed by the former include but are not limited to the physical facts presupposed by the latter. Thus, the argument is that, to the extent that we are constrained by four-dimensional physics to explain qualia, this would either result in physicalism in the form of eliminative materialism, since, as I will argue, we cannot deduce qualia based on four-dimensional physics, or we must posit a dualism, as Jackson ends up doing, or otherwise reach for some extra-scientific form of explanation like panpsychism which posits that the mental is fundamental and ubiquitous, to explain qualia. As such, I argue that, for physical descriptions to a priori entail qualia, a viable physicalism must entail Higher Dimensional Physicalism to deduce qualia. This will help explain, among other things, the private nature of phenomenal consciousness.

Further, according to Higher Dimensional Physicalism, the features of mind picked out by qualia, understood as private and accessible only from the first-person perspective, will thus be regarded as instantiated in and constituted by the higher physical dimensions conjectured by string theory/M-Theory, beyond those picked out by four-dimensional physics. Thus, the mind is understood as extending into the empirically unobservable, physical dimensions conjectured in modern physics, as discussed. This is how it can be understood to explain the private nature of phenomenal consciousness since the conjectured physical dimensions are themselves beyond empirical observation (at least by current and foreseeable technologies for measurement). On the philosophical level of analysis, Higher Dimensional Physicalism, then, can show how qualia fits in with research on the neurobiological level of analysis, as well as with physical facts of the natural world in general, based on natural principles.

There are, however, other attributes of phenomenal consciousness besides its private nature, including subjectivity and its qualitative or phenomenal aspect. In general, Qualia denote experiential states, meaning that their referent are *in practice* only accessible in the a posteriori, from the first-person perspective, sometimes described as the "what-it´s-like" property of phenomenal consciousness (Nagel, 1974). Importantly, since my thesis argument is based on a theory-based physicalism, the argument is that it is possible to provide physical description of facts about the target phenomenon ´consciousness´, by a priori entailment. The a priori entailment will resolve the epistemic gap from the third-person perspective, to arrive at knowledge of facts about qualia as physical facts only *in principle.* This, however, does not provide knowledge of experiential states *in practice.*

In other words, no amount of theoretical or physical description, i.e. a priori entailment, about the phenomenal property of redness will reveal what it is like to experience

perceptions of red, any more than exhaustive scientific knowledge of a bat's mode of perception via echolocation will tell us what it is like for the bat to have perceptions via this modality (Nagel, 1974). As such, while my thesis argument will aim to show that we can have a priori entailment of physical facts of qualia *in principle*, this will not yield insights into experiential states *in practice*.

This, however, does not represent an epistemic gap between facts of qualia and physical facts in the sense that Jackson sought to demonstrate by his knowledge argument, however. Jackson sought to demonstrate how qualia cannot be entailed by physical facts even in principle, as will be developed in chapter 1, which is what I will aim to refute. Thus, I will also leave to the side such issues as inverted spectra, where the question is raised whether the phenomenal property of redness you see when you look at a tomato would correspond to what I see when looking at lush grass and vice versa (See *section 4.11 Phenomenal Concepts and Privacy* in Papineau, 2004). The resolution of this issue, whether there exists variation between the privately accessible phenomenal properties between different people, such that the content-specific colour valence differs between people from the first-person perspective, has no real bearing on whether the existence of such qualia can be described based on physical facts. This is the central claim made by Jackson, that the phenomenal property of any given qualia, irrespective of the specific nature of any given quale, exists outside physical facts. This claim is what I am to refute.

This broader point about qualia and their relationship to physical facts is what is at stake in Jackson´s Knowledge argument against physicalism. This epistemic gap discussed above led Jackson to argue for an ontological gap between physical facts and phenomenal properties, as with dualism. This dualism can take on different forms, including interactionism, overdetermination or epiphenomenalism.

Interactionism submits there must be some psychophysical law governing a two-way causal relation between physical facts and qualia. Overdetermination submits that for every physical cause there is always a concurrent mental cause. Epiphenomenalism submits there must be a psychophysical law governing a one-way causal relation, whereby qualia are rendered causally impotent or inert, resulting as an epiphenomenal byproduct of the causal relations between physical properties, events, or processes.

In all these cases, qualia are relegated to an extra-physical ontological domain, existent outside the world we perceive and describe by physics, thus left dangling in some unnamed metaphysical space. This would mean there is an epistemic gap between qualia and

physical facts not just *in practice* as I have defined it, but *in principle* and on a metaphysical and ontological level.

In Jackson's article, "Epiphenomenal Qualia" (1982), where he first introduces what has become standardly referred to as "the knowledge argument", because of his knowledge argument which he considers having refuted physicalism, he goes on to argue for dualism in the form of epiphenomenalism to account for qualia. This view of qualia as epiphenomenal with respect to physical facts of nature renders qualia causally impotent while, on the level of ontology, submits a form of dualism, as just described. This implies a much stronger claim about the epistemic gap than what is necessary or desirable since it introduces an ontic divide between the two. The resolution of the epistemic gap, which must now be thought to include some kind of psychophysical law(s), thus becomes less tractable since qualia and physical facts become separated and bound by knowledge about different epistemic domains on the level of ontology, as per dualism.

There are two main reasons I consider this ontological gap to be undesirable. The first is that it is less ontologically parsimonious since it submits a dualism. By Ockham´s Razor, it is better to have one ontological domain than two, given that it is sufficient to provide a priori entailment of the target phenomenon. The second reason is that it confounds consciousness with a greater sense of mystery, since the features of mind picked out by qualia become understood to exist in some otherworldly domain which remains otherwise opaque and, epistemically, beyond, or at least far more difficult for, our ability to comprehend and describe. Indeed, this last point about the comprehensibility of qualia in principle becomes an important point for Jackson when he lays out his argument for why these are to be understood as epiphenomenal. I will address this in chapter 3.

Now, elsewhere in Jackson (1999) he has described what he would consider to be a viable physicalism. He puts it as follows,

> "Thus, one way materialists can show that the psychological has a place in their world view is by showing that the psychological story is entailed by the story about the world told in the materialists´ favored terms. We will see, however, that it is not just one way; it is the only way." (1999, p. 484.)

Materialism is here treated as synonymous with physicalism and thereby used interchangeably. Moreover, the psychological story referred to here would include features of the mind like those picked out by qualia. Presumably, the favoured terms by physicalists would be those based on physical facts, treated as exhaustive to describe the target

phenomenon of consciousness. This is precisely what I will be arguing for by my theory-based physicalism, to provide a priori entailment of qualia. After this challenge posed, he then goes on to explain how a supposed incompleteness of physicalism would depend on the relevant notion of supervenience, where the point of contention would be whether there can be independent variation between qualia and physical facts.

What does this mean? Minimal physicalism is often defined as reliant on supervenience (Stoljar, 2021; Chalmers, 2016; Hansen, 2010.) Expressed in simplified terms, in the context of philosophy of mind, supervenience states that there can be no change in consciousness without physical change (Lewis, 1999). So, if qualia can vary independently of physical facts, this would refute physicalism. We will unpack this in chapter 1 when discussing the Knowledge Argument in some detail, before assessing some common objections raised by Lewis and Churchland. In the following chapter, I will argue that these standard objections are themselves insufficient to refute the knowledge argument since they are bound by what I call Four-Dimensional Physicalism, as previously discussed.

Before we get to the details of the knowledge argument, however, allow me to first make clear the general structure of my own argument. According to the conclusion of my argument, the target phenomenon ´consciousness´ becomes understood as a multidimensional spatiotemporal entity, i.e., a physical system or process. I argue that consciousness can be entailed a priori, in accordance with a theory-based physicalism as discussed above, meaning we can deduce a priori all facts of consciousness, including qualia, to provide exhaustive description of the target phenomenon in principle.

On this view, this entity, our consciousness, becomes understood as a physical system instantiated in and constituted by physical dimensions beyond the four dimensions of space and time we interpret based on empirical observation, as discussed above. This is the main point of my thesis argument and serves as the basis for the theory-based physicalism to be tried, which is considered as conditional on the theoretical frameworks introduced by string theory or M-Theory in modern physics.

I will explain the relevant features of string theory in chapter 3. For now, we can briefly note that string theory was developed as a unifying theory of fundamental reality to bring together the four natural forces as part of a complete and unified theory in physics. Thus, the gravitational force becomes unified with the other three natural forces described in Quantum Field Theory, namely the electromagnetic force and the strong and weak nuclear forces, as well as with the elementary particles described by the Standard Model. There are currently three superstring theoretical models on offer and two heterotic string theoretical

models, each including six additional spatial dimensions beyond the three we can observe empirically. M-Theory unifies these five string theoretical models and includes one further spatial dimension, for a total of ten spatial dimensions and one temporal dimension. My argument supposes that qualia, taken as physical facts in this way, extend into these physical dimensions, to explain its private nature and why it remains unobservable from a third-person perspective.

Further, I will suppose that the perceived divide between *mind* and *body,* between *qualia* and *physical facts*, can be characterised as conceptually distinguishing between *phenomenal* and *physical properties* of consciousness, respectively. As Papineau (2004, p. 5) points out, this view of ontological monism and conceptual dualism is by now a standard materialist view.[3]

By phenomenal properties I mean that which is picked out by the qualia Jackson discusses. Starting from a rough description, these are the properties of consciousness we have access to only in private, from a first-person perspective, distinctive as that subjective sense of *what-it's-like* discussed by Nagel (1974).

On a more specific characterisation of phenomenal properties, I follow Block (1995) in his useful distinction of Phenomenal-Consciousness from Access-Consciousness. Thus, I consider phenomenal properties to be understood as possibly non-representational, non-intentional, non-functional properties. These more fine-pointed characterisations will have bearing on my response to the objection raised against the knowledge argument by Lewis and Nemirow, as we will see in subsection 1.3.

As mentioned, the phenomenal properties are conceptually distinguishable from the physical properties of consciousness. The latter are here treated in terms of the neurobiological processes picked out by the neural correlates of consciousness (NCC). Thus, I will treat the physical properties associated with consciousness as those constituted by the empirically observable neurobiological structures, events, and processes in the organism, most notably the brain, which remain empirically observable by standard scientific observation and measurement. For another handle for the reader to connect to with my sense of physical properties, these include what are picked out by the "easy problems of consciousness", as formulated by Chalmers (1995). These are the functional properties implicated in the neurobiological mechanisms and processes supporting cognition, content-

[3] To support this claim, he refers to a whole slew of philosophers before him who bear this out, including Peacocke (1989), Loar (1990), Sturgeon (1994), Hill (1997), Hill and Mclaughlin (1998), and Tye (1999). I may add P. F. Strawson (1974) who also discussed this possibility.

specific perceptions, and functional consciousness in general. More on this in chapter 2, as previously alluded to.

Moreover, unlike the phenomenal properties, the physical properties associated with consciousness as so construed are not private since they obtain in the four dimensions of space and time apt for empirical observation. Thus, we can observe and measure these neural correlates, these physical properties, in a laboratory setting, from a third-person perspective. This is why, when I refer to these physical properties as picked out by the NCC, we can speak of a *Four-Dimensional Physicalism* for short*,* since these are subject to the dimensions of physical space and time we have access to by empirical observation in general.

These physical properties, then, are of course spatiotemporally constituted, but are constrained by what I call Four-Dimensional Physicalism, which remains amenable to observation in four dimensions of space and time and in accordance with classical physics. Thus, if we treat the brain and body as a classical system we can, in principle, gain a complete and exhaustive description of the physical properties of consciousness, as so construed.

However, to the extent that the brain is treated as a classical system to explain consciousness, thus constrained by Four-Dimensional Physicalism, this cannot provide a priori entailment of qualia since, by the private nature of phenomenal properties, these are understood as constituted in the higher dimensions of physical space which are themselves beyond empirical observation. This point will be developed later in chapter 4 where I discuss the plausibility of the thesis in general, by providing an argument for why the neural correlates of consciousness should be thought to include quantum mechanical principles as a minimum for my thesis argument to work.[4]

Accordingly, to the extent that physicalism is constrained by classical physics or four-dimensional physics as so construed, I argue that Jackson´s argument is successful. This is so since the physical properties as defined above do not include the higher dimensions of space conjectured in string theoretical physics, which is where I argue phenomenal properties are instantiated and constituted.

So, to state it in full, I argue that the knowledge argument holds as a refutation against Four-Dimensional Physicalism but not against Higher Dimensional Physicalism, which takes physics to include the added spatial dimensions conjectured by string theory or M-Theory. We

[4] The reason for this, as will be discussed, is the quantum mechanical principles governing string theoretical properties in general.

can compare, then, a reformulation of Jackson´s knowledge argument provided by Chalmers (2003) to a basic formulation of what I will be arguing for. Chalmers formulates the knowledge argument against physicalism thus,

> P1: There are truths about consciousness that are not deducible from physical truths.
> P2: If there are truths about consciousness that are not deducible from physical truths, then physicalism is false.
> C: Physicalism is false. (2003, p. 7)

I will aim to refute Jackson´s knowledge argument by undermining its first premise which, as Jackson (1999) has emphasised, is taken to include any complete physics, current or future. Compare this to my argument, which can be stated according to four premises, resulting in five conclusions which may be deduced.

> P1: If and only if there are truths about consciousness which are not deducible from any physical truths, the knowledge argument is successful.
> P2: There are truths about consciousness that are not deducible from physical truths constrained by Four-Dimensional Physicalism.
> P3: All truths about consciousness are entailed and deducible from physical truths constrained by higher-dimensional physicalism.
> P4: If Mary knows all truths about consciousness that are deducible from physical truths constrained by higher-dimensional physicalism, she knows all truths about consciousness in principle.
> What I consider the main conclusions to be drawn from this is:
> C1: If Mary only knows truths about four-dimensional physics, she does not know all the truths about consciousness.
> C2: If physical truths are understood as constrained by Four-Dimensional Physicalism, the knowledge argument is successful.
> C3: If Mary only knows truths about Four-Dimensional Physicalism, she does not know all the physical truths.
> C4: If Mary knows all the physical truths, she knows all the truths about consciousness in principle.
> C5: The knowledge argument is unsuccessful.

Recall how Higher Dimensional Physicalism as I have defined it is presupposes and entails Four-Dimensional Physicalism, along with the conjectured spatial dimensions of string theory/M-Theory or some similar multidimensional theory to obtain in physics. The inclusion of these higher dimensions is motivated by the need to provide a priori entailment of qualia, to explain the basis for the private nature of phenomenal properties and how they fit in with brain process and the natural world in general, describable in terms of physical facts.

Thus, these conclusions work when applied to knowledge to be deduced in the relevant sense. This is otherwise referred to as a third-person account (Churchland, 1999), forthcoming, or we can say that it constitutes knowledge in principle, not in practice. To the extent that the epistemic gap argued by Jackson turns on a kind of epistemic gap which is understood to imply an ontological gap between qualia and physical facts in principle, if I am considered successful, the argument will be refuted by the main argument of my thesis.

In chapter 1, I will begin by addressing Jackson´s knowledge against physicalism and consider some previous objections raised by Lewis and Churchland, specifically. To the extent that these are insufficient to reject the argument in full, or lack further detail to provide a positive account, this can be shown to motivate my overall thesis argument.

In chapter 2, I will consider the limitations of Four-Dimensional Physicalism, treated in terms of the scientific investigations on consciousness by observing the NCC. I argue that these are inadequate to provide entailment of phenomenal properties of consciousness since they only report on Access-Consciousness states as developed by Block (1995)

In chapter 3, I provide a brief description of string theory/M-Theory and develop the basis for what I call Higher Dimensional Physicalism. Here I will look at how this argument fits in with some previous philosophical work, especially that of Sjöstedt-Hughes (2022) and John Searle (1999). Further, I will contrast this with the epiphenomenalism Jackson went on to defend as a consequence of his Knowledge Argument against physicalism.

In chapter 4, I will address Jackson´s subsequent argument for dualism in the form of epiphenomenalism. I will examine in turn each of the objections against epiphenomenalism he raises to defend against, to assess the efficacy of the conclusion he draws. I will then provide some contradistinctions with the Higher Dimensional Physicalism developed in the preceding chapters.

Finally, I will supply some concluding remarks. Let us first examine Jackson´s Knowledge Argument against physicalism.

Chapter 1: Physicalism and The Knowledge Argument

In this chapter we will begin by addressing the Knowledge Argument

1.1. The Knowledge Argument

I will now explain in more detail Frank Jackson´s knowledge argument (Jackson, 1982) to test how my thesis can yield a new perspective on its efficacy. The knowledge argument is presented as a refutation of physicalism, to demonstrate how physicalism cannot account for phenomenal properties, or qualia, and is thus considered insufficient as a philosophical theory of consciousness. Jackson provides two imaginative examples to spell out his argument. The first one concerns Fred who can discriminate another colour the rest of us are unable to perceive (us neurotypical chromatic creatures would be unable to distinguish it from red). So, the argument goes, regardless of how much physical information we can learn about the neurological intricacies Fred´s unique visual system, we can never learn about Fred´s experience of the colour we are unable to discern. I will not dwell too much on this example since the same point is enhanced by his second imaginative example which seems to have become paradigmatic for the knowledge argument in the wider literature, perhaps because it seems more effective. I will devote my attention to this second example for the remainder of the discussion.

In the second example, we are asked to imagine Mary, a brilliant scientist who has been confined to a room wherein her visual experience of the world has been in monochrome for the duration of her life. She has learned all about colour experience *via* a black-and-white television screen (and perhaps from books printed on gray paper). Mary has a perfect and complete understanding of all the physical facts about colour perception. Upon her release from the room, when she encounters a ripe tomato, she is astonished by her experience of its phenomenal property red. The question Jackson poses is, has she learned something new, a new phenomenal fact about what it's like to see red?

To this question, Jackson would respond in the affirmative. Jackson´s argument is devised to demonstrate that before her release from this room, Mary does not know all the facts, since she does not know facts about the phenomenal property associated with qualia, in this case phenomenal redness. Thus, the argument goes, facts about the phenomenal property associated with qualia must be understood as outside or beyond what is entailed by physical facts about colour perception. This is so since, for all her knowledge of the physical facts,

these do not entail the phenomenal property attributable to qualia, the facts about which she can only learn by direct observation. As he puts it,

> "Will she *learn* anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had *all* the physical information. *Ergo* there is more to have than that [physical information], and physicalism is false." (Jackson, 1982, p. 130.)

Thus, he concludes, physical facts are inexhaustive in explaining qualia and physicalism is false.

We can express the argument according to a formulation based on two premises leading to a conclusion against physicalism:

(1) Mary knows all the physical facts.
(2) Mary does not know all the facts.
(3) The physical facts do not exhaust all the facts. (Chalmers, 2003)

Chalmers (2003) has also offered the more general reformulation of the argument as stated in the introduction:

(1) There are truths about consciousness that are not deducible from physical truths.
(2) If there are truths about consciousness that are not deducible from physical truths, then physicalism is false.
(3) Physicalism is false.[5] (2003, p. 7)

On its face, there are two main ways to refute the knowledge argument to invalidate its conclusion. First is to reject (1), that there are truths about consciousness which are not deducible from physical truths. Second is to reject (2), there are truths about consciousness which are not deducible from physical facts, then physicalism is false. As we have seen, I will target the first premise. Before we get there, let us look Lewis´s reformulation of the knowledge argument, to enhance our interrogation of its efficacy.

[5] In the original formulation, Chalmers uses materialism in place of physicalism, but these terms are treated as synonymous and thereby used interchangeably. For the sake of continuity, in my reformulation I make use of the term physicalism.

1.2. Lewis´s Reformulation of the Knowledge Argument

Before raising his objection, Lewis formulates a strongman version of the knowledge argument based on the premise which he calls "the Hypothesis of Phenomenal Information" (1999, p. 583). He formulates this premise to dispense with "lookalike arguments" which may mischaracterise what is picked out by Jackson´s knowledge argument, and thus fail to refute it on the appropriate grounds. I also consider Lewis´ reformulation of the knowledge argument helpful, so I will take some time to bear it out in full, as well as inject an important distinction on a subtle point about the issue of inverted spectra alluded to in the introduction.

While Jackson provides a short and concise formulation in his original article, along the lines discussed above, there the argument is cast in terms of the limitations of physical information. We might say that Lewis inverts the argument by invoking what he calls "The Hypothesis of Phenomenal Information", to the same effect as the original knowledge argument and its conclusion against physicalism. Thus, the Hypothesis of Phenomenal Information Lewis goes on to develop is a restatement of Jackson´s initial hypothesis, to show that phenomenal properties associated with qualia exist independently of physical information. As Lewis puts it,

> "No amount of the physical information that black-and-white Mary gathers could help her know what it was like to see colors … There is a natural and tempting explanation of why physical information does not help. That is the hypothesis that besides the physical information, there is an irreducibly different kind of information to be had: *phenomenal information.* The two are independent. Two possible cases might be exactly alike physically, yet different phenomenally." (1999, p. 583.)

To the first sentence, both Lewis and I would agree. It is true that you cannot know what it is like to experience the phenomenal property of redness or the taste of vegemite before you have in fact had that experience, any more than we can know exactly what it is like to experience the navigation of time and space by way of echolocation, as a bat does (Nagel, 1974). This goes to my claim that we can provide a priori entailment of qualia based on physical facts in principle, even while these physical facts cannot in themselves reveal what it is like to experience phenomenal properties in practice.

Lewis goes on to develop what he takes the Hypothesis of Phenomenal Information to mean, before applying it in his objection against Jackson´s knowledge argument. The Hypothesis of Phenomenal Information is defined in terms of its capacity for *elimination of possibilities*. He suggests that, although information of physical facts narrows down the

physical possibilities, it could be seen that "they leave open a range of phenomenal possibilities," since these are understood as independent. In this way, when "we acquire phenomenal information, possibilities [of phenomenal information] previously open are eliminated; and that is what the experience is like." (Lewis, 1999, p. 583).

To help make this intelligible in less abstract terms, I may concretise it by way of an example of my own. Mary (before her release) has all the physical facts about colour perception, *ex hypothesi,* as we have seen. She knows that what we describe as "red" corresponds to that which on the spectrum of visible light is on the wavelength of about 700nm, blue corresponds to that on the wavelength of about 450nm (if my memory serves from physics class, Mary will know better).

This knowledge of the physical facts about electromagnetic radiation *narrows down* the possibilities of physical information considerably, eliminating so many other possibilities (it might have been that what we perceive as red would be located on some other wavelength of electromagnetic radiation or otherwise according to some different physical laws altogether). She may wish to pontificate on this and suggest there are various hues of red, corresponding to a slightly wider range between the wavelengths of 700-630nm, and so forth, but the possibility range of this physical information about redness on the electromagnetic spectrum has eliminated so many other possibilities of physical information, vastly constrained as it is by these tightly fixed parameters. The same goes for physical information in terms of all possible neurophysiological brain states from which she could infer colour perception.

To continue with my example, for all this physical information, if Mary is released from the room and enters another well-lit room painted in all manner of brilliant colours, without access to relevant technologies for measurement of brain states or electromagnetic radiation, there would be no way for her to determine the colours of the paint based on this physical information alone. Thus, all the possibilities for the relevant *phenomenal information* remain open. The full range of possibilities for phenomenal information remain open until her trusted friend and colleague, Fred, enters the room, points to one section of the painted wall and states, "that is what we call blue." This eliminates the other possibilities on the possibility range of *phenomenal information* for Mary. She now learns to denote the relevant colour percept, the phenomenal information, to that which she now knows to be blue (independently of the physical information about the same).

To Fred´s amazement, he learns from Mary that what he denotes as blue corresponds to the wavelength of about 450nm on the electromagnetic spectrum of visible light, and so

they both come away from the experience having eliminated possibilities of physical and phenomenal information, both learning something new in the process. (They then go on to share a meal of vegemite on toast, to experience that taste for the first time, mostly out of spite and to ruin a good example for Lewis.)

To provide a second example of my own to illustrate Lewis´s point, in even simpler terms we might say that a child, knowing nothing about electromagnetic radiation or the rods and cones of your retinal ganglion cells and the intricate processes of the fronto-parietal network and occipital lobe in the brain, can easily identify and discriminate between blue toys and red toys alike (no doubt a source of much envy to our brilliant scientist, Mary, who devoted her life to understanding how all this works). The child achieves this without recourse to any knowledge about the physical information, but based on the phenomenal information alone.

Allow me to digress briefly here, to reiterate a point raised in the first section of this chapter concerning issues of inverted spectra, as Papineau (2004) discusses in the context of phenomenal concepts and privacy (and indeed brushes to the side). To the point here, whether the phenomenal property associated with the qualia known by Fred from the first-person perspective as "blue" seems different from the "blue" Mary learns of from her first-person perspective as he points it out to her, where, in this case, the relevant phenomenal property for both is associated with what the physical information tells us occurs on the wavelength of about 450nm (i.e., the physical information remains the same), this is beside the point. If Mary consistently perceives "red" under normal circumstances wherein Fred consistently perceives "blue" from their first-person perspectives, based on perceptions of the same physical object (e.g., a smurf or a tram in Oslo) about which their phenomenal information is obtained, this is not what the knowledge argument is concerned with. The concern for Jackson, as he himself is careful to point out, is that facts about the phenomenal property itself which is associated with qualia, irrespective of the content of any given quale, exists independently from any facts about the physical property or information about the same.

This is a subtle but important point which he draws out at some length in §III "The "What it is like to be" Argument". Here, he discusses Nagel´s (1974) point about our inability to imagine a bat´s modality for sonar perception. Jackson extricates the phenomenal *property* of any given experience from our ability to extrapolate based on knowledge of other experiences (in the case of bats, the experiences we would extrapolate from would necessarily need to be based on different modalities, since we do not possess the modality for sonar perception). So, he says,

> "When I complained that all the physical knowledge about Fred was not enough to tell us what his special colour experience was like, I was not complaining that we weren´t finding out what it is like to *be* Fred. I was complaining that there is something *about* his experience, a property of it, of which we were left ignorant." (Jackson, 1982, p. 132.)

And a little further down page, towards the conclusion of this section of the article,

> "Nagel argues that the trouble with bats *et al.* is that they are too unlike us. It is hard to see an objection to physicalism here. Physicalism makes no special claims about the imaginative or extrapolative powers of human beings, and it is hard to see why it need do so.
>
> Anyway, our knowledge argument makes no assumptions on this point. If physicalism were true, enough physical information about Fred would obviate any need to extrapolate or to perform special feats of imagination or understanding in order to know all about his special colour experience. *The information would already be in our possession.* But clearly it isn't. That was the nub of the argument." (Jackson, 1982, p. 132.)

So, while we can imagine that your red is different from my red, such that I am left to extrapolate based on my experience of blue what would be your experience of red, this is not to the point. Jackson´s concern is knowledge about the phenomenal property itself, of which, according to Jackson, we are left ignorant by physical information alone.

Now, some hard-headed anti-physicalists might wish to argue that the determination of inverted spectra lies beyond what any given physical facts can tell us about the specific phenomenal property of any given quale as it seems from the first-person perspective. Indeed, this is also part of what I mean when I argue that physical descriptions can provide a priori entailment in *principle,* not in *practice.*

Moreover, to a hard-headed anti-physicalist, I might retort that it would remain equally indeterminable also based on knowledge of phenomenal facts in themselves, because of the private nature of first-person experience as discussed in the first section. Thus, these privately accessible phenomenal facts cannot in themselves tell us about the same experience from the first-person perspective of other conscious minds.[6] As such, the resolution of this issue would pose no more of a challenge to the physicalist than it would for a dualist, to settle the phenomenal facts about any given quale, as it seems to different people or conscious creatures in general. Either way, I do not believe this goes to the intended point Jackson wished to make, which sought to demonstrate how phenomenal properties themselves are

[6] I will leave to the side here the philosophical problem surrounding "other minds" as construed in terms of the conceivability argument and concerning philosophical zombies (Chalmers, 2003), which would need to be treated separately. The focus here is the knowledge argument against physicalism. However, the issue of knowledge about qualia of *other* minds is central also to the knowledge argument, as will become apparent in the forthcoming discussion of objections raised by Lewis and Churchland.

pbeyond physical description, not just in practice but in principle, and so I will leave it to the side.

Thus, to return to Lewis, the key point he raises which is to be identified here is how phenomenal information is understood as *independent* from physical information. As Lewis explains, this independence between physical information and phenomenal information, between physical facts about properties of physical description and phenomenal properties in themselves, I might add for the sake of continuity and further clarification on the subtle nuance outlined above, "provides a powerful argument to refute any materialist of the mind." (1999, p. 585).

Indeed, Lewis goes on to explain how showing that phenomenal information is independent of physical information would refute not simply one form of physicalism or other but, if successful, would represent a knockdown argument of minimal physicalism. This, then, would obviously also extend to my theory-based physicalist thesis. This is because physicalism is standardly understood to presuppose a supervenience thesis as a minimum (Lewis, 1999. Stoljar, 2021. Chalmers, 2016. Hansen, 2010).

To say that F supervenes on G is to say that that there can be no change in a set of properties F without a difference in a set of properties G. In the context of philosophy of mind, then, if physicalism is right, there can be no difference in consciousness without a difference in physical properties. The basis for this physicalist claim, however, is precisely what is undermined by showing that phenomenal information is independent of physical information, as Lewis points out, which is what Jackson purported to demonstrate by way of his knowledge argument, leaving physicalism to wither on the vine.

Having reformulated Jackson´s knowledge argument against physicalism in terms of the Hypothesis of Phenomenal Information as discussed above, Lewis draws his conclusion. "The knowledge argument works. There is no way to grant the Hypothesis of Phenomenal Information and still uphold materialism. Therefore I deny the hypothesis." (1999, p. 586.) As such, Lewis concedes that he cannot refute it outright. Rather, he argues that there are reasons to see how this hypothesis should seem more peculiar than one might first suspect. Mainly, this is because we could take the Phenomenal Information Hypothesis to refute more than just physicalism but take it to apply also to parapsychological phenomena. Paul Churchland (1999), forthcoming, points out the same as we will see. First, let us consider Lewis´s objection.

1.3. Lewis´s Objection

Lewis´s objection begins with pointing out three peculiarities about the knowledge argument, cast in terms of the Hypothesis of Phenomenal Information, to make it seem less appealing as a refutation of physicalism. The first reason for its peculiarity which Lewis points out goes to the same point I raised in my discussion of the inverted spectra, namely our inability to determine unknown phenomenal properties in practice from the first-person perspective. According to the Knowledge Argument, this is due to us "[thinking all along that] physical information was inadequate to explain the phenomena of mind." (Lewis, 1999, p. 589.) However, Lewis points out how, as I mentioned above, this presents no more of an issue for physicalists than it does for dualists. He bears this out by demonstrating how Mary would be no more poised to know phenomenal properties in practice given a complete and perfect understanding of "parapsychological information", including lessons in all manner of dualistic psychophysical laws governing the instantiation of phenomenal properties, since such lessons do not reveal anything about the phenomenal nature of the experience of the phenomenal property associated with qualia.

This leads to the second reason, along the same lines. Since Lewis treats the knowledge argument in terms of elimination of possibilities about phenomenal information, the presentation of qualia as principally ineffable, as Jackson posits, "resists treatment" for acquiring knowledge about them not just according to physical or parapsychological information, but indeed on any given principle. Thus, it remains inaccessible to be deduced by any logic or mathematics.

> "Phenomenal information cannot be logical or mathematical, because lessons in logic and mathematics no more teach us what a new experience is like than lessons in physics or parapsychology. When someone doesn´t know what it´s like to have an experience, where are the alternative open possibilities? I cannot present to myself in thought what it might be like to taste Vegemite." (1999, p. 589.)

In fairness to Jackson, again, to me this does not seem so strange since I maintain that phenomenal properties or experiential states cannot be entailed in practice in the a priori (indeed, by definition *a priori* would mean "prior to experience"). However, I do not believe this supports Jackson´s claim which treats phenomenal properties as beyond physical description even in principle. This is the claim I aim to refute. The issue this seems to raise for Lewis, however, and as we will see, is that it resists representation in thought.

This leads to the third peculiar reason Lewis points out, which targets Jackson´s subsequent argument for epiphenomenalism to explain qualia. What Lewis reacts to here, and

on this point, I agree, is that such an understanding of qualia as epiphenomenal supposes that "it is strangely isolated from all other sorts of information; and this is so regardless of whether the mind works on physical or parapsychological principles." (1999, p.589.) As I put it in the first section of this chapter, this leaves qualia dangling in some unnamed metaphysical space; it becomes ontologically separate and distinct from the world we act in an perceive as part of normal life. This is indeed a strange turn which, to me, seems to be nothing but a relic of dualism inherited from the modern period starting with Descartes; it presupposes these preconceptions dualism has wrought. I will return to this point in chapter 4 where I provide some contradistinctions between my thesis argument and Jackson´s epiphenomenalism. If I am successful, we will see how this leap by Jackson is not necessary.

First, let us review Lewis´s objection. Because of these peculiarities borne out by the knowledge argument, or the Hypothesis of Phenomenal Information as Lewis dubbed it, Lewis thinks it prudent to replace the hypothesis with another which he would consider less spurious. He calls it "The Ability Hypothesis." As he concedes, it is indeed essential to reject the former hypothesis in favour of the latter, to save materialism. The Ability Hypothesis is based on what Laurence Nemirow (1980) suggests learning a new experience is like. Lewis supplies a lengthy quote to bear out the relevant point, it starts as follows,

> "Some modes of understanding consist, not in the grasping of facts, but in the acquisition of abilities… As for understanding any experience, we may construe that as an ability to place oneself, at will, in a state representative of the experience." (Quoted in Lewis, 1999, p. 591.)

As such, we can see how Lewis will go on to develop a specific type of representationalism to explain the target phenomenon, based on certain abilities to form intentional states about experience, that is, to represent the experience *at will*, along the lines Nemirow describes.

Lewis goes on to suggest that Jackson confuses and equivocates on two different types of knowledge, knowledge-that and knowledge-how. Knowledge-that refers to the sort of knowledge gained by lessons about facts, whereas knowledge-how refers to the acquisition of abilities. More specifically, knowledge-that refers to the sort of knowledge you can acquire by lessons in physics, parapsychology, or (to supply an example of my own) watching instructional videos on YouTube when you wish to learn how to sew a button. Knowledge-how would refer to the acquisition of the relevant abilities to perform a task, like riding a bicycle, wiggling your ears, or eating food using chopsticks.

Now, not all abilities are understood as a form of knowledge, like the ability of strength or to provide sufficient funds to achieve some ends, as are Lewis´s examples. However, the relevant abilities you gain when you have a new experience, according to

Lewis, are the abilities to *remember*, to *imagine*, and to *recognise*. Roughly, we can equate knowledge-that to what I refer to as knowledge in principle. The knowledge-how, however, deviates from what I refer to as knowledge in practice since, as I defined it, knowledge in practice does not rely on any ability to interpret or classify phenomenal properties according to some theoretical or conceptual scheme.

While knowledge-that may aid you in the acquisition of these abilities since you receive lessons on how to perform a task, like receiving descriptions of the relevant geometrical shapes and dimensions of a C-38 locomotive to help you recognise it by sight, Lewis argues it often does not contribute enough to your knowledge-how. Thus, as Lewis explains, "This is why music students have to practice." (1999, p. 593.) To further concretise the distinction, when you taste Vegemite for the first time, you acquire the ability to later *remember* what that experience was like. If you have the same experience again, and in a non-obvious context like if I surprise you with a closed sandwich where the topping is concealed from view, you have acquired the ability to *recognise* it as Vegemite once you taste it, based on previous experience of the same taste. Moreover, if you later come to forget exactly the quality of what that experience was like, you can "very likely retain your ability to *imagine* such an experience." (1999, p. 592.)

The nub of Lewis´s distinction about these two forms of knowledge is to turn away from the discussion about knowledge understood in terms of "information", along the lines of that discussed in the previous section, both in terms of physical and phenomenal information alike. Indeed, as Lewis puts it, "if the Ability Hypothesis is the correct analysis of knowing what an experience is like, phenomenal information is an illusion." (1999, p. 593.) Thus, the phenomenal properties associated with qualia that Jackson discusses, the phenomenal information as Lewis refers to it as, is displaced in favour of cognitive abilities to remember, recognise, and imagine.

To my mind, there is something about Lewis´s Ability Hypothesis which does not really capture or entail the basis for phenomenal consciousness. For one thing, based on this analysis whereby phenomenal information is to be regarded as an illusion, there is a sense in which the Ability Hypothesis may come close to eliminative materialism in the form of illusionism. I will not dwell too much on this point, since I will not have space in this paper to provide a detailed argument against illusionism. I will simply supply a brief remark based on Keith Frankish´s (2016) account of illusionism as a philosophical theory of consciousness, to point out in rough terms how I regard it as deficient.

On Frankish´s account, the notion of "phenomenal properties" becomes replaced by "quasi-phenomenal properties" and "intrinsic subjectivity" becomes replaced by "inherent subjectivity", among other such distinctions. The aim seems to be to invalidate or ameliorate the apparent irreducibility of phenomenal properties in physical facts. This would make the target phenomenon ´consciousness´ seem more tractable for continued scientific investigation, not to get mired in phenomenology. However, there does not seem to be much substantive difference in the role these revised terms have for our apparent phenomenal consciousness, beyond a presupposition that such revisions renders phenomenal consciousness illusory. The aim, as far as I can understand, is to displace both substance and property dualism, by way of eliminative materialism. Hence Frankish´s call to replace "the Hard Problem of Consciousness" with "The Illusion Problem" (Frankish, 2016).

For my part, like Jackson and perhaps most philosophers of mind, I too am somewhat of a "qualia freak" and believe we are hard-pressed to deny the reality of phenomenal consciousness, understood in terms of phenomenal properties (which Lewis believes must be displaced). Indeed, I would maintain that these "properties", whatever they are, is the very basis of our experience. Thus, I do follow Jackson in saying that there are such phenomenal properties associated with qualia, what Lewis deemed beyond the materialist's ability to reject outright.

Whether Lewis would subscribe to Frankish´s call to replace The Hard Problem with The Illusion Problem, I cannot know. This is a more recent development, after Lewis´s time. However, what Frankish and the illusionists are doing amounts to what I would call a "gerrymandering" of the relevant terms, to purport a resolution of the issue, though it is hard to see how it carries much weight. I will leave this here as my stated opinion in place of a more careful analysis, which will have to be left to the side.

To the extent that Lewis does not go so far as illusionism on this form, his account can be seen to rely on a kind of representationalism about phenomenal consciousness. Representationalism was first and foremost a central concern in discussions concerning intentionality. Only later has it been introduced also more widely to philosophy of mind, to provide a theory of consciousness as such. As Lycan and Pautz explains representationalism,

> "Indeed, there are now multiple representational theories of consciousness, corresponding to different uses of the term "conscious," each attempting to explain the corresponding phenomenon in terms of representation. More cautiously, each theory attempts to explain its target phenomenon in terms of *intentionality*, and assumes that intentionality is representation." (Lycan, Pautz, 2020)

As such, relying on representational abilities to explain the phenomenon of consciousness is to invoke the possibility of forming intentional states, or to represent in thought experiential states *at will,* as Lewis and Nemirow supplies. This representational and intentional aboutness is the basis for the Ability Hypothesis, including the abilities to remember, imagine, or recognise the relevant experiential state, to represent them in thought. However, as Jackson has also pointed out in a subsequent rebuttal to objections raised against the knowledge argument, including the basis for the Ability Hypothesis,

> "Powers of imagination is not to the point. The contention about Mary is not that, despite her fantastic grasp of neurophysiology and everything else physical, she *could not imagine* what it is like to sense red; it is that, as a matter of fact, she *would not know."* (Jackson, 1999, p. 567.)

Indeed, and to give Jackson his due, I would agree that this is not what his argument points out.

Now, recall how Lewis does not pretend to refute the knowledge argument; he has already conceded that he cannot refute it outright. Rather, he rejects the argument in favour of the Ability Hypothesis. However, I would agree with Jackson that the central concern about the phenomenal property associated with qualia would remain unaddressed by such a move. Accordingly, the Ability Hypothesis does not entail the relevant phenomenal property of qualia I would subscribe to. Allow me to elaborate, and to show why I believe it is inadequate to explain the target phenomenon in full.

The notion of know-how, which would displace the phenomenal properties discussed by Jackson does not seem to capture, or entail, the relevant experiential quality associated with phenomenal consciousness which I consider indispensable. It is fine to say that when we experience for the first time a new sensation, it enables the acquisition of abilities to remember, imagine, and recognize it subsequently. In this way, we can represent the experiential state in thought after the fact, at will. This representational account makes it easier to see how we can form intentional states about experiential states.[7] However, what is that *thisness* about which we acquire these abilities to represent after the fact? What is it that we learn to recognise, and what is *this* that we can remember or imagine later? We can represent it to ourselves in thought along these lines, according to such abilities, but what is it that these representational abilities pick out, if not the ability to re-present some property associated with the qualia itself? Thus, the Ability Hypothesis turns on a form of

[7] Leave aside here also any concerns about the feasibility of freedom of will and what that might mean.

representationalism about phenomenal consciousness without addressing the aboutness of the experiential state they refer to.

This characterisation of Lewis as maintaining representationalism is a standard interpretation and seems uncontroversial as far as I understand.[8] However, I would concede to Jackson that such representationalism offered by the Ability Hypothesis and its analysis of the target phenomenon does not much help us "qualia freaks", as Jackson had styled it. As I stated in the first section, I consider the phenomenal properties Jackson points out as possibly non-representational, non-intentional, and non-functional. More specifically, I follow Block (1995) in his distinction between Phenomenal-Consciousness and Access-Consciousness. I believe this will be useful to identify the inadequacy of Lewis´s Ability Hypothesis and its given analysis.

1.4. Ned Block´s Phenomenal Consciousness

In his article, ´On a Confusion About a Function of Consciousness,´ Block (1995) comments on the vagueness of the term ´consciousness´, suggesting that it is "a hybrid, or better, a mongrel concept" (1995). He suggests that, as a term, it denotes various, subtly different phenomena which become conflated in ways we do not always fully recognise. According to Block, this has a detrimental effect on our reasoning about the target phenomenon.

With P-Consciousness, Block is quick to admit his inability to provide a non-circular definition, suggesting that he can merely point to it. This is not an uncommon observation.[9] To increase the accuracy of our pointing, he fixes the term with some further synonymous description. He explains how he takes P-Consciousness to be related to or otherwise characterised by experiential properties. We can understand this as the phenomenal properties Jackson picks out by his knowledge argument, as discussed here.

According to Block, such properties may indeed sometimes be representational, but not necessarily. This is to say that phenomenal consciousness does not necessarily form an intentional belief. Thus, Block defines P-Consciousness as distinct from cognitive, intentional, or functional properties in general. In this way, the phenomenal properties associated with qualia do not necessarily gear toward the know-how pointed out by Lewis,

[8] Indeed, he is mentioned in the Stanford Encyclopaedia Entry on representationalism cited above.
[9] See for example Mandler, 1975. George Miller is another who put it succinctly enough by saying, "Turning a tool on itself may be as futile as trying to soar off the ground by a tug at one´s bootstraps." (Quoted in Güzeldere, 1999, p. 24.) A knife cannot cut itself, as I have once heard it said.
Güzeldere has referred to this circularity as the "recursive impossibility" of identifying the basis of our experiential states (1999, p. 24.)

which concerns our ability to use chopsticks, recognise a C-38 locomotive by sight, play music, and so the abilities to remember, imagine, and recognise in a functional mode.

Block contrasts Phenomenal-Consciousness with Access-Consciousness. He defines A-Consciousness as the aspect of consciousness used to promote rational control, described as a "cluster concept" consisting of three parts:

1) It is inferentially promiscuous, meaning, poised for use as a premise in reasoning.
2) It is poised for rational control of action.
3) It is poised for rational control of speech. (Block, 1995)

Further, Block characterises A-Consciousness as necessarily representational, intentional, functional, and thus involved in executive system functioning. Thus, A-Consciousness represents the various modes of cognition and information processing we rely on to enact rational control - to determine thought, speech, and behaviour. This would be the know-how the Ability Hypothesis describes.

With these definitions in view, it becomes plain to see how Lewis´s Ability Hypothesis entails only Access-Consciousness, without addressing Phenomenal-Consciousness as so construed. The invocation of knowledge-how as discussed above relies on representational and functional notions *qua* A-Consciousness, since know-how revolves around the acquisition of cognitive abilities including imagining, recognising, and remembering previous experiential states. All these abilities require the formation of intentional states about the relevant experiential state, to conjure these at will, as Lewis puts it, to represent them in thought subsequently. This is presented as obviating the need for the phenomenal properties Jackson discusses, which Lewis admits he cannot refute outright.

Yet, to qualia freaks such as Jackson and me, this representational analysis given by the Ability Hypothesis leaves out the possibly non-representational, non-intentional, non-functional quality of P-Consciousness defined by Block; the *thisness* about which these representational abilities take for their basis. Like Jackson and Block, I too would consider this aspect of phenomenal consciousness indispensable to provide a complete explanation of the target phenomenon.

In sum, in the analysis given by the Ability Hypothesis, the "phenomenal information", the phenomenal property associated with qualia identified in Jackson´s knowledge argument remains unaddressed by Lewis´s account. For this reason, given the definitions provided by Block and to which I subscribe, the Ability Hypothesis Lewis promotes remains insufficient to explain all aspects of consciousness.

Moreover, it cannot be seen as a viable alternative to the knowledge argument, or the Hypothesis of Phenomenal Information, to provide a complete account of the target phenomenon unless we subscribe to some form of eliminative materialist view about phenomenal consciousness, which I would reject. Thus, having considered the efficacy of Lewis´s objection, I conclude that the challenge posed by Jackson remains and physicalism still hangs in the balance.

1.5. Churchland´s Objections

The context in which Churchland raises his objections against the knowledge argument is in his 1985 paper, ´Reduction, Qualia, and the Direct Introspection of Brain States´. The main topic and theme of the discussion in this paper is intra- and inter-theoretic reducibility in science. On the topic of phenomenal properties, then, the concern, as Jackson sought to demonstrate, is the irreducibility of qualia.[10] In his discussion of the Knowledge Argument, the central points raised by Churchland´s objections to go against it may seem similar to those raised by Lewis, as has been discussed above.

In the first instance, like Lewis, Churchland complains that Jackson equivocates on the notion of knowledge. He objects that, though both premises in Jackson´s knowledge argument rely on an apparently transparent notion of "knows about," the senses of "knows about" are not univocal as they are expressed and exploited in the basic premises found in Jackson. The tightened version of Jackson´s Knowledge Argument is put thus by Churchland,

> (1) Mary knows everything there is to know about brain states and their properties.
> (2) It is not the case that Mary knows everything there is to know about sensations and their properties.
> Therefore, by Leibniz's law,
> (3) Sensations and their properties =/= brain states and their properties (Churchland, 1985, p. 23)

Further, he also comments on Lewis´s Ability Hypothesis and states that,

> "Lewis and Nemirow plump for the "ability" analysis of the relevant sense of 'knows about', but they need not be so narrowly committed, and the complaint of equivocation need not be so narrowly based." (1985, p. 23.)

[10] This irreducibility is precisely the basis for «The Hard Problem of Consciousness" as Chalmers (1995) has identified it.

I suggest that raising the objection in the broader sense as Churchland would have it does makes it more plausible, because it would not directly imply or necessitate a stronger form of representationalism as that which the analysis given by the Ability Hypothesis reveals. As I have just argued, this stronger sense of representationalism leaves out the *thisness* about which it represents. I will pick up this point momentarily but allow me to slow down a little.

Churchland suggests that the relevant senses in which Mary "knows about" brain states and properties stated in (1) can be referred to as propositional knowledge. He calls this "knowledge by description". This knowledge by description is what I have been referring to as "knowledge in principle", what Lewis referred to as Knowledge-that. The relevant sense in which Mary "knows about" sensations and their properties stated in (2) can be referred to as "knowledge by acquaintance". I take this knowledge by acquaintance as consistent with what I have been referring to as "knowledge in practice". However, it is broader than what Lewis submits with his knowledge-how, such that, by my estimation at least, it need not be committed to the kind of representationalism committed to by Lewis, as discussed above.

Now, Churchland argues that when we resolve this equivocation between different ways of "knows about" which Jackson has been tacitly and implicitly committed to, the two stated premises do not entail the conclusion drawn by Leibniz´s Law, which Churchland identifies as the relevant logical principles by which the Knowledge Argument is understood to work.

Leibniz´s Law concerns the principles known as the Identity of Indiscernibles and its converse, the Indiscernibility of Identicals. A standard formulation of the former goes like this,

- "if, for every property *F*, object *x* has *F* if and only if object *y* has *F*, then *x* is identical to *y*." (Forrest, 2010.)

The latter would be the converse principle, I might formulate it as follows.

- if object *x* is identical to object *y*, then every property *F* of object *x* is materially equivalent to every property *F* of object *y.*

As Forrest tells us, "Sometimes the conjunction of both principles, rather than the Principle [of Identity of Indiscernibles] by itself, is known as Leibniz's Law." (2010)

Thus, in the tightened version of Jackson´s Knowledge Argument as formulated by Churchland above, if the properties picked out by brain states in (1) are substantially or extensively different from, i.e., materially inequivalent to, the properties picked out by sensations in (2), then the two would be non-identical, according to Leibniz´s Law. What Churchland targets here is the scope and content of what the relevant forms of knowledge

invoked in the first and second premises of the argument entail, to show how the properties these pick out refer to the same.

Consequently, Churchland´s analysis seeks to demonstrate that these are materially equivalent, only that the types of knowledge about the same are different. He goes on to suggest that the type of knowledge Mary learns when becoming acquainted with phenomenal redness as in (2) "seems to be a matter of having a representation of redness in some prelinguistic or sublinguistic medium of representation." (1985, p. 17.) He shows how when we fit this as the type of knowledge Mary lacks in (2), the conclusion that redness sensation does not equal brain states is no longer entailed by Leibniz´s law. Granted, this type of knowledge via a prelinguistic or sublinguistic medium of representation is not given to her by knowledge of description, but this is not to say that it is somehow beyond physicalist assumptions. As he puts it,

> "The materialist can freely admit that one has "knowledge" of one's sensations in a way that is independent of the scientific theories one has learned. This does not mean that sensations are beyond the reach of physical science. *It just means that the brain uses more modes and media of representation than the simple storage of sentences"* (1985, p. 24, italics in original quote)

Like we have discussed in the previous sections via issues of inverted spectra and what Lewis pointed out with Mary´s lessons on parapsychology, Churchland reiterates this same point. He shows how the very same argument provided by Jackson would hold if Mary was a brilliant "ectoplasmologist", meaning she knew all the about "the ectoplasmic processes underlying vision". Given a complete and perfect propositional knowledge within this supposed ectoplasmology, Mary would still not be able to know what it is like to see red, for precisely the same reason, identified as Jackson´s equivocation on the ways in which Mary "knows about" the facts. As Churchland proclaims, "Dualism is therefore inadequate to account for all mental phenomena!". (1985, p.25)

In this way, Jackson could be said to "prove too much". Moreover, understood in this way what the Knowledge Argument would prove is certainly more than what Jackson had intended to show. This is evident by the fact that Jackson goes on to develop an argument for epiphenomenal qualia, which admits a form of dualism, based on his conclusion by the Knowledge Argument taken as going against physicalism. As both Lewis and Churchland object, however, this would fail to obtain for the reasons specified above.

Now, the second shortcoming of the knowledge argument Churchland supplies is that it fails to fully appreciate how powerful the imaginative capacity of Mary would be if she really had a "utopian" understanding of all relevant neuroscientific information at her

disposal. While he concedes that "direct deducibility" of phenomenal states would perhaps be impossible, he also suggests that this is too strong of a demand on reduction in the relevant sense.

What he does argue is that Mary would likely be able to *imagine* with a high degree of accuracy what phenomenal redness, for instance, would seem like due to the highly sophisticated introspective apprehension of relevant brain states her exhaustive neuroscientific knowledge confers. Indeed, Churchland argues that it does not seem impossible to him that Mary would come very close to imagining with significant accuracy what such a state of experience might seem like, even in the absence of any external stimuli to produce the corresponding brain state. He suggests that,

> "None of these philosophers [who would be swayed by the Knowledge Argument] have even begun to consider the changes in our introspective apprehension of our internal states that could follow upon a wholesale revision in our conceptual framework for our internal states." (1985, p. 25**.)**

How plausible this seems; I am not entirely sure. Whether Mary could make such inferences about phenomenal properties, to visualise phenomenal redness by her imaginative prowess, based on introspective grounds alone (in the absence of the relevant external stimuli, as Churchland explicitly states) and without ever having experienced colour perception, as Churchland would have us believe, seems a little far-fetched. Granted, I am not as brilliant as Mary, so perhaps I shall suspend my judgment. Churchland goes on to concretise this further by way of some examples concerning Mary´s ability to discriminate colour sensations since these are "structured sets of elements". He suggests we can test the efficacy of her introspective apprehension in the same way we can test an expert musician´s ability to identify the relevant set of chords played backed to him from an array of different samples, based on his examination of the sheet music (which is not an uncommon ability).

It will be helpful here to bring in some further clarification provided by Jackson in his rebuttal of Churchland´s objections, however. We can recall how, as I stated in section 1.3., Jackson stated that powers of imagination are not to the point. In the same paper, entitled, *´What Mary Didn´t Know´* (1999) Jackson also states the following,

> "The knowledge Mary lacked [before her release] which is of particular point for the knowledge argument against physicalism is *knowledge about the experience of others,* not about her own. When she is let out, she has new experiences, color experiences she has never had before. It is not, therefore, an objection to physicalism that she learns *something* on being let out. Before she was let out, she could not have known facts about experience of red, for there were no such facts to know. That physicalists and nonphysicalists alike can agree on … The trouble for physicalism is that, after Mary sees her first ripe tomato, she will realize how

> impoverished her conception of the mental life of *others* has been *all along.*" (1999, pp. 568-9.)

This distinction about how the central concern for Jackson is Mary´s knowledge of the mental life of *others,* not her capacity for introspective apprehension, is interesting and will prove pertinent to my overall thesis argument. The basis for my argument is, as mentioned in the first section of this chapter, that the phenomenal properties of qualia, the basis of our experiential states, are instantiated in and constituted by higher spatial dimensions. Thus, for Mary to have sufficient knowledge about the mental life of others, she will need to have sufficient knowledge of the relevant string theoretical physics to include qualia. I will develop this in the following chapters.

First, we can note briefly how Churchland (1999) responds to this rebuttal. He here reiterates the same objection concerning Leibniz´s Law. I agree that this still stands, and that the issue with the Knowledge Argument is an equivocation on the ways in which Mary "knows about" the facts. However, he further specifies by way of some examples, by recalling the distinction made by Lewis concerning knowledge-that and knowledge-how. His example concerns the a competent golfer´s ability to execute a golf swing based on the relevant relevant *motor* representation, perhaps identified in his cerebellum, perhaps in his motor cortex. (1999, p. 572) However, as I have argued, this would remain insufficient to explain the target phenomenon since this would constitute Access-Consciousness, not Phenomenal-Consciousness, as per the preceding discussion in section 1.4.

Further, we can recall the quote by Jackson (1999) referenced above in section 1. As he puts it, a viable materialism would need to show how the psychological features belongs to their view by "showing that the psychological story is entailed by the story about the world told in the materialists´ favoured terms." (Jackson, 1999). We may note here how this kind of "entailment" referred to by Jackson would qualify as a kind of reduction, which is what he sought to demonstrate by the Knowledge Argument was impossible based on the irreducibility of qualia. This alleged irreducibility is what Churchland will challenge.

First, Churchland asserts that,

> "There is no reason we must be bound by the crude divisions of our prescientific idioms when we attempt to give a precise and positive explication of the equivocation in Jackson´s argument." (Churchland, 1999, p. 572.)

On this basis, he goes on to provide a detailed neurobiological description of "creatures with trichromatic vision (i.e., with three types of retinal cones)." (1999). He then proceeds to

explain how the coding of colour information occurs on the level of the brain. The details are not too important. The important point is how this propositional knowledge, the *knowledge by description* Churchland (1985) defined it as, what I call knowledge in principle, does indeed entail the non-discursive knowledge Mary would come to know about upon observing phenomenal redness. To me, it is difficult to see how this does not qualify as an entailment of the psychological features according to materialist terms, as Jackson previously specified would be the necessary requirement for a viable materialism, as we have seen.

Indeed, in his final objection, Churchland goes so far as to suggest that the equivocation on ways of "knows about" as we have discussed simply begs the question against physicalism. If Mary is in possession of knowledge about all the propositional facts concerning colour perception, the assertion in premise (2) simply presupposes that "sensory qualia form a metaphysically distinct class of phenomena beyond the scope of physical science," (1999, p. 575) by exploiting the equivocation on forms of knowledge about the fact as we have discussed.

I am inclined to agree with this assessment of the knowledge argument as question begging. I think Churchland put it succinctly in his final remarks when suggesting that the resultant view Jackson arrives at reveals more than anything the "ideological grip" whose clutch we struggle to shake.

> "[It is an assumption that subjective experience is] metaphysically distinct from the objective physical properties addressed by orthodox science. It is not a surprise, then, on this view, that one might know all physical facts, and yet be ignorant of some domain of these nonphysical qualia. The contrast between what is known and what is not known simply reflects an antecedent metaphysical division in the furniture of the world." (1999, p. 575.)

This antecedent metaphysical division baked into premise (2) is what would reveal the *petitio* which reveals the alleged question begging of the argument as such. This antecedent metaphysical division might also become clearer in the discussion concerning Jackson´s argument for epiphenomenalism later in chapter 3. For now, another main point I wish to take away from Churchland is among his concluding remarks in the final section of his (1999) response to Jackson, on "Converting a Third-Person Account into a First-Person Account."

The key point to be understood, as far as I am concerned, is the sense in which reducibility of qualia in terms of what I refer to as a priori entailment does not make any claims to reveal the content and of phenomenal redness as it would seem from the privileged view of the privately accessible first-person perspective. However, this does not "signify anything about their metaphysical status", as Churchland does well to point out.

To reiterate, the continued investigation into the neurobiological bases for the modes of representations of qualia which were previously discursively inarticulable can reduce the target phenomenon in terms of a priori entailment. This is the view I will argue to defend in my thesis argument too, and what I mean when I suppose that we can provide knowledge in principle about facts of qualia, even while knowledge in practice necessarily involves us having the experiential state. The absence of knowledge about the latter does not, however, lead us to suppose an ontological division between qualia and physical properties in general, as with dualism. This is simply a non sequitur stemming from an equivocation about the different ways in which Mary is said to know about the facts, as Churchland describes it.

1.6. Conclusions

There is something fishy about the Knowledge Argument, as Lewis and Churchland both point out, since its conclusion is effective not just against physicalism, but it would also blow dualism out of the water in the same blast. If it is the case that qualia are inherently irreducible on any principle, as both Lewis and Churchland have shown would be the case if we are to accept the Knowledge Argument at face value, not only can there not be any science of consciousness, but there can also not be any substantive or propositional philosophical insights concerning the basis for phenomenal consciousness whatsoever.

In this case, the basis for this phenomenological experience would otherwise remain totally ineffable and inscrutable to us. Thus, this basis would be the equivalent of a black hole to any other mode of knowledge acquisition, beyond our private experience of it as it seems to us. As such, any information about qualia, beyond the imminent sensation it arouses in us as part of our experience, to continue the metaphor with black holes, would remain past the event horizon.

If this is the case, it would indeed force philosophers to either dispel the notion of phenomenological properties associated with consciousness by submitting to eliminative materialism, as with illusionism, or we must confess that we simply can never truly understand what consciousness is and how it relates to Nature, and how it arises from the self-organising neurobiological processes of our bodily organism. Given this understanding of what the Knowledge Argument can be said to show, “The Hard Problem of

Consciousness" (Chalmers, 1995) would seem all too optimistic in its formulation, and it would have to be redubbed "The Impossible Problem of Consciousness."[11]

However, this total ineffability of qualia as so construed was obviously not what Jackson took the Knowledge Argument to show. This is evident by the fact that he went on to defend a form of dualism in the form of epiphenomenalism, thinking he had undermined only physicalism. Like both Lewis and Churchland, I do believe it would be premature to consign to epistemic darkness the prospect of a theoretical description for consciousness, even while I would maintain the reality of phenomenal properties as an indispensable aspect of the target phenomenon. Lewis, as we have seen, rejected the validity of phenomenal information owing to the inherent intractability such a construal seemed to represent to him. For this reason, I consider his objection against the Knowledge Argument insufficient to entail the target phenomenon. The representationalism he offers in its place can only explain Access-Consciousness, not the full extent of our phenomenology to include phenomenal properties which are possibly non-representational, non-intentional, and non-functional.

Churchland does not plump for such an outright rejection, at least not explicitly. However, what I would point out here is that what we have been discussing so far, the paradigmatic example of phenomenal properties of qualia Jackson picked out by his Knowledge Argument, relies on visual perception. It is important to note how the visual system is a *modality* of consciousness, not consciousness itself.

Accordingly, in Churchland´s discussion concerning the phenomenal properties associated with qualia of colour perception, a question we might ask is how this relates to Mary´s knowledge about the structure and availability of phenomenal properties associated with consciousness states in general? How can Mary know about the occurrence of such phenomenal properties as experienced by other minds, as Jackson is careful to point out?

This is what I will now address in the next chapter, where I will argue that the general structure and availability of phenomenal properties associated with qualia cannot be located or fully identified in four dimensions of space and time. The ensuing discussion will concern the neurobiological research surrounding both content specific NCC and the prospect of a full NCC.

[11] Granted, I can agree that the latter is a distinct possibility. Indeed, like most students of philosophy, perhaps, I often amuse myself in my attempts to grasp the hidden depths of my own ignorance. Perhaps we are no more apt to grasp the basis of our own consciousness than the scientist slugs Jackson goes on to discuss are apt to understand the nature of the wider universe based on their observations from the point of view of their domicile on the seabed (I will return to this point in chapter 4).

Chapter 2: Four-dimensional Physicalism and the Neural Correlates of Consciousness

As described in the introduction, the thesis argued for here is a theory-based physicalism. As such, what we refer to as the phenomenal properties of consciousness can be understood as physical properties themselves, if physical properties are understood to include those properties entailed by the extra spatial dimensions string theory introduces. Thus, this will be an argument for what I call Higher-Dimensional Physicalism. By contrast, what I call Four-Dimensional Physicalism, is constrained by any physics describing nature according to four dimensions of space and time, as with classical physics. These are the spatio-temporal dimensions amenable to empirical observation, and the standard interpretation for continued investigations into the neural correlates of consciousness (NCC) to explain the target phenomenon. As I will argue, these NCC only entail Access-Consciousness as discussed in the previous chapter, not Phenomenal-Consciousness.

Thus, when I speak of physical properties related to consciousness, broadly speaking what I take this to refer to is the various electro-chemical signalling throughout the organism, most notably the brain. Moreover, since most philosophers of mind or consciousness researchers in general probably do not consider string theory when talking about physical properties, I will regard the "physical properties" discussed here as those constrained by Dour-dimensional physicalism as defined in the introduction. On the neurobiological level of analysis, these physical properties constitute the neurophysiological structures, mechanisms, and events researchers observe while charting the NCC.

What I will aim to demonstrate in this chapter is the basis for my claim that the physical properties observed and measured on the neurobiological level of analysis invariably exist in the standard four dimensions of space and time amenable to empirical observation, representing only Access-Consciousness. As such, to the extent that Mary has knowledge of physical facts constrained by what I previously defined as "Four-Dimensional Physicalism", I will concede to Jackson that the Knowledge argument holds as a refutation against such a physicalism.

2.1. The Neural Correlates of Consciousness

We can begin with what are called content-specific NCC. These are standardly defined as "the minimum neuronal mechanisms jointly sufficient for any one specific conscious percept."(Koch, Massimini, *et al.,* 2016) These content-specific NCC, then, pick out neural

correlates of experiential states related to some modality of consciousness, as in the discussion of Mary where the paradigmatic example became that of the visual system, concerning the colour perception of redness, specifically.

It is important to note how the determination of the relevant neuronal mechanisms for consciousness also relies on a subjective measure. They require a participating subject to report on the experience they are currently undergoing as we make note of the corresponding neuronal activity occurring simultaneously. As a result, there is no obvious, scientifically rigorous way to provide theoretical structure for the subjective experience of consciousness according to a no-report paradigm. As Koch et al. explains, "in a clinical setting, simple behavioural criteria are often used to infer consciousness, such as the ability to respond to a command."(2016, p, 307) This is typically done by way of verbal reports or by pushing a button in response to given stimuli. In this way researchers can develop quantitative indices associated with conscious states.

Our observations of the relevant neuronal mechanisms are augmented by the utilisation of various technological instruments for measurement like functional magnetic resonance imaging (fMRI), electroencephalography (EEG), or magnetoencephalography (MEG). Moreover, we can induce or manipulate conscious states by utilising optogenetic or transcranial magnetic stimulation (TMS) techniques. Nevertheless, the claim here is that these physical properties, according to the standard usage of the term, while providing descriptions for NCC, they can never really spill the beans on the structure of the phenomenal experience itself.

While neuroscientists can map the relevant metabolic brain activity or neuronal mechanisms associated with various conscious content based on the reported experience of participants, the inference of conscious content will necessarily remain reliant on behavioural responses of the participating subject. Even in so called no-report paradigms for NCC, the inference of conscious awareness is drawn based on eye movement or specific brain activity indicative of conscious engagement based quantitative indices derived from the growing corpus of previously retrieved data of the NCC (Overgaard, 2017).

Although such no-report paradigms are without a doubt highly useful in a clinical setting, to determine whether a person is in a minimally conscious state or suffering from locked-in syndrome, for example, arguably, these still constitute a kind of behavioural response. Again, this is because they are predicated on inferences based on eye movements or relevant quantitative indices previously discovered to correlate with conscious experience.

However, and this is the claim, this does not yield any further insight into the structure of the phenomenal property of consciousness itself.

To elaborate on this, we may say that the conscious content of visual stimuli is associated with specific activations of the human visual system. Distal stimuli are encoded in the retinal ganglion cells in the eyes, before being passed through the optic chiasm for further processing primarily in the occipital lobe.[12] Thus, when presented with an image of a face or even while asked to imagine a face, we come to expect that the same neuronal networks be engaged in each case. In the case of visual experience these would be identified with specific activity in the temporo-parietal-occipital region. However, arguably this does not provide any real insight into the phenomenal nature of sight itself or otherwise show how the phenomenal property of consciousness might be structured. This is because, while vision is an integral part of conscious experience in typical human subjects, it is a modality of consciousness, not consciousness itself *qua* phenomenal awareness.

The same can be said for any other modality of consciousness. Take speech-processing for another example. We may come to understood it as somehow related to the structure of the brain known as Broca´s area in communication with the temporal cortex and the motor cortex, to effectuate speech. Though this might provide certain quantitative indices revealing insights into the relevant mechanisms informing cognition, in no way does it relate the experience of the subject itself. Put differently, even if we managed to achieve a complete account for content specific NCC, the phenomenal property of consciousness and its relationship with the physical substrate of the brain inevitably remains obscure by this approach.

This is another way to look at the Knowledge Argument, then, since what I am suggesting is not the epistemic gap based on an equivocation on the ways in which Mary knows about the facts, as previously discussed. Rather, the claim here is that what these neural correlates pick out are those connected to Access-Consciousness as discussed in chapter 1. This is because these correlates are based on the verbal reporting or behavioural responses (even with the rational control of eye-movements) of the participating subjects. In

[12] For a more detailed description, see for example Crick and Koch (1998) Here, the authors discuss not so much the basis for consciousness as such, rather, they provide an account for the content-specific mechanisation of a given modality of consciousness, namely visual processing.
Victor Lamme (2004) has also argued for the phenomenal awareness of visual stimuli based on recurrent processing in the brain. Interestingly, he provides what might be characterised as a completely mechanistic description based on feedback-feedforward processing occurrent between the striate and extrastriate cortices. Yet, he argues that the realisation of recurrent processing across different cortical structures supports realism on the phenomenal property of consciousness.

this way, even a complete description of the relevant physical and cognitive mechanisms for content-specific, conscious experience when perceiving the colour red are insufficient to capture the *qualia* (i.e., the phenomenal property) of the resulting experience of redness, by this method. Thus, this does not provide entailment of the non-representational, non-intentional, non-functional characteristics of Phenomenal-Consciousness.

This leads us to the next point, which is the prospect of a full NCC. Full NCC is understood as "the neural substrates supporting conscious experiences in their entirety, irrespective of their specific contents." (Koch, Massimini, *et al.,* 2016, p. 308) Can a complete account of the neural substrates supporting consciousness on the neurobiological level of analysis provide entailment of phenomenal properties of consciousness in general? Put differently, if Mary had complete and perfect understanding of a full NCC, could she know about the structure and availability of qualia in other conscious minds?

As neurobiologist, Dick Swaab, states, "the cerebral cortex and the thalamus are crucial for consciousness – as is a functional link between the two." (Swaab, 2015, p. 159.) Incidentally, the functional link he identifies between the two are working neural connections in the fronto-parietal network. A more detail-specific account can be found in Dendritic Integration Theory (DIT) (Aru, Larkum, 2020). In my opinion based on what I have seen, DIT could be considered the most specific explanation currently available to provide a truly fundamental neurobiological theory of consciousness.

What this research demonstrates is an experimentally verifiable mechanism to support consciousness by the brain. The specific mechanism is identified as the coupling, or signal matching, of segregated data input signal streams propagated globally in the brain. This broadcasting of the relevant signal streams is referred to here as "reverberating activity in thalamocortical loops" (2020, p. 819.). As Aru and Larkum point out, this is consistent with other prominent theories such as Global Neuronal Workspace Theory (GNWT) (DeHeane, *et al.,* 2011; Mashour, *et al.,* 2020) and Integrated Information Theory (IIT) (Tononi, Koch, 2015). The claim that DIT can be considered unifying of GNWT and IIT is also remarkable, since the two are sometimes considered as competing theories.

According to DIT, the coupling of these signals occurs in Layer 5 Pyramidal cells (L5P cells), specifically. Thus, it "proposes that the central operation underlying consciousness is the integration between apical and basal compartments of L5P cells." (Aru, Larkum, 2020, p. 815). L5P cells are complex dendritic cells constituting cortical columns which connect the thalamus with the cerebral cortex (as is crucial for consciousness, as

Swaab reminds us in the abovementioned). These cells may be abstracted in such a way as to divide them into three compartments. These compartments include,

1) the apical compartment whose dendrites arborizes in the cerebral cortex,

2) the coupling compartment in the middle of the cell axon, and

3) the basal compartment whose dendrites arborizes in the thalamus (see figure 3.)

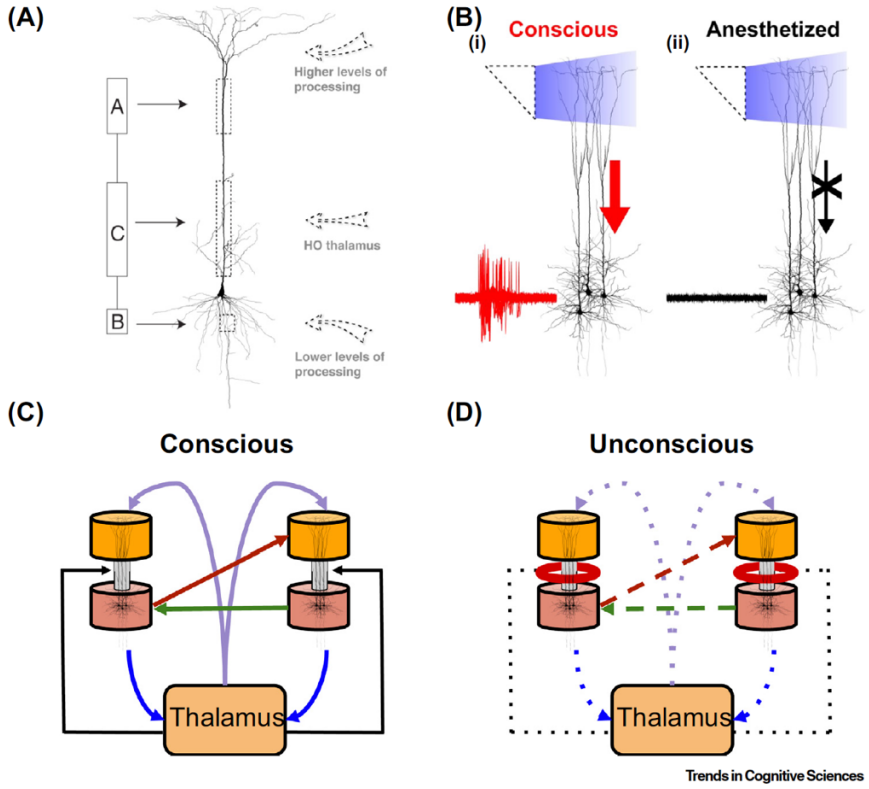

Figure 3: Diagram of relevant Layer 5 pyramidal cells. In (A) box ´A´ points to the apical compartment whose input stream is associated with higher levels of processing, box ´B´ points to the basal compartment whose input stream is associated with lower levels of processing, and box ´C´ points to the coupling zone mediating signals between the segregated input streams of the apical and basal compartments. In conscious states, successful mediation by the coupling zone is present as shown in (B) and (C). When the coupling zone is inhibited, such as under the influence of effective anaesthetics, this is demonstrated to result in unconscious states, also shown in (B) and (C).[13]

As the authors explain,

> "The basal compartment integrates feature-specific information [such as that provided by feedforwarded input signals retrieved from processed sensory data], whereas the apical compartment receives internal variables that relate to, or could be associated, with the feature. The internal variables could be context and expectations, but they could also be semantic knowledge or episodic memory about the feature." (Aru, Larkum, 2020, p. 821).

Consciousness occurs when there is a matching signal between these two segregated data input streams, as mediated by the coupling zone. Further, when communication via the

[13] This diagram is taken from (Aru, Larkum, 2020, p. 817).

coupling zone is inhibited, as happens under general anaesthesia, consciousness ceases. This is true even when stimulating activation of the apical compartment using optogenetic or transcranial magnetic stimulation (TMS) techniques.

No doubt, this is a truly remarkable advancement, providing description of the specific neuronal mechanism implicated in supporting consciousness by the brain. If correct, dendritic integration of segregated data input streams might indeed be the defining characteristic of consciousness, as Aru and Larkum propose. On a more philosophical level of analysis in connection with the discussion in chapter 1, this raises an interesting point about the kind of representationalism raised by the Ability Hypothesis discussed in section 1.3 and 1.4. The representational abilities to remember, recognise and imagine as part of "knowledge-how" *qua* Lewis, can be seen as that which provides cognitively accessible context and expectations, corresponding to the pre-established semantic knowledge or episodic memory these "internal variables" received at the apical compartment refer to. The salient *thisness* I asked about, however, could be seen to correspond to the "feature specific" information received by some sensory data input or other afferent nervous signalling integrated at the basal compartment of the L5P cells.

Further, we can understand this input data at the basal compartment as the pre-linguistic and sublinguistic mode of representation as Churchland referred to it (which would represent the *thisness* about which we recognise, remember, and imagine as part of an intentional state after the fact). Conscious states arise only when there are matching signals between these two segregated data input streams, identified and mediated at the coupling zone, as we have seen. In this way, we can see how the sort of representational abilities entailed by know-how (tantamount to the internal variables integrated in the apical compartment) are necessary but not sufficient for consciousness states, since it fails to include the salient *thisness* by which these cognitive abilities are previously informed and with which they would correspond as you receive new input data of the type integrated at the basal compartment.

Moreover, there is another more general claim made by the authors of DIT, relevant to our current discussion. To quote, "more than that, we contend that the biophysical mechanisms within the pyramidal neurons do conceptual work in terms of understanding consciousness." (Aru, Larkum, 2020, p. 820). On a certain interpretation, and on the neurobiological level of analysis, this may be accurate. I will not spend much space unpacking all the relevant details of their subsequent arguments. The basis for this claim by the authors is that DIT is both consistent with, and unifying of, other prominent

neuroscientific strategies and theories, like GNWT, IIT, and methods of predictive coding, which are sometimes regarded as at odds with each other in specific ways.[14] Thus, the unification of these theories indicates a more fundamental neurobiological theory of consciousness. Though this may be true, on a more philosophical level of analysis, what does it tell us about the phenomenal property of consciousness?

Suppose for the sake of argument that DIT represents a final account of full NCC. Put differently, suppose we have discovered the necessary and sufficient neurobiological conditions to provide complete explanation for conscious awareness. To return to the question as I phrased it earlier, given the complete and perfect knowledge of the full NCC in this way, would this help Mary know about the structure and availability of qualia in conscious minds in general? I would suggest that it clearly would not.

For one thing, I would suggest that what the quantitative indices resulting from our investigations into the NCC reveal, both with content-specific and full NCC, are functional processes associated with Access-Consciousness, as defined in the previous chapter. To the extent that these rely on verbal and other behavioural reports, and even in cases when they do not as with monitoring comatose patients or the like, they show the concurrent neural activity associated with functional states of the participating subject, to effectuate rational control.

Secondly, to make this point clear, consider notion of multiple realisability (Bickle, 2020). This introduces the prospect of other conscious systems which may include other biological organisms equipped with a neural substrate (or some equivalent information integration structures) substantially different from the neural architecture in the human case (like plants, octopi, or people of the Andromeda galaxy) or otherwise silicon-based forms of artificial intelligence. Would the available data supplied by the NCC help Mary deduce or identify phenomenal properties in other conscious systems? Could she know about this qualia of *other minds,* as Jackson put it? Let me also specify that what I am asking refers to Mary´s knowledge about phenomenal properties understood in terms of knowledge in principle, to provide a priori entailment on a third-personal account.

As just indicated, I would suggest that she quite clearly could not. The reason being that a complete account of full NCC in the human case, in large part based on the subjective measure we ourselves supply as previously discussed, cannot possibly determine whether

[14] The most prominent implementations of predictive coding the authors of DIT cite are Friston (2005) and Bastos, *et al.* (2012).
See also, Hohwy and Seth (2020). Interestingly, Hohwy and Seth do not present their strategy for predictive processing as a theory for consciousness, rather as a framework for the systematic study of consciousness in a rigorous, neuroscientific mode.

another system includes the phenomenal property of consciousness. Indeed, as I have suggested, it only entails Access-Consciousness. To help demonstrate this, we might consider how Mary might determine consciousness in another system as mentioned above.

2.2. Consciousness in other systems: Conscious Machines

Let us take as our example the prospect of Artificial General Intelligence (AGI) or Artificial Consciousness (AC). Already, there are large language models (LLMs) predicated on certain machine learning programmes capable of passing the Turing-test. The Turing-test was a kind of thought experiment introduced by Alan Turing as early as in 1950 to address the question of whether machines can think. (Turing, 1950) It consists of an "imitation game" whereby an interrogator is tasked with distinguishing between participant A and B, based on the given replies to the interrogator´s questioning. One of the participants is a human subject and the other a machine, defined as a digital computer more specifically. To the extent that the interrogator is unable to determine which is the digital computer, it could be seen that "thinking machines" have been realised, given the definitions Turing supply.

Of particular interest is the fourth objection to this thought experiment Turing considers in his discussion, namely "The Argument from Consciousness." (Turing, 1950, pp. 445-457) Here he quotes Professor Jefferson´s Lister Oration for 1949.[15] Part of the full quote Turing supplies is as follows,

> "Not until a machine can write a sonnet, or compose a concerto *because of thoughts and emotions felt* [my italics], and not by the chance fall of symbols, we agree that machine equals brain – that is, not only write it but know that it had written it." (Turing, 1950, p. 445.)

The aspect of consciousness Jefferson sought to establish, and isolate, is that "what-it´s-like" property of consciousness later discussed by Nagel (1999). Thus, on the conception of conscious machines, Jefferson raises the bar by adding the qualifying condition that the generation of their outputs must be accompanied by a subject of experience with the capacity for phenomenal awareness about the composition of its creative activity, if they are to be considered truly conscious. Following Ned Block´s useful distinction as discussed in the previous chapter, I would suggest that the phenomenal property associated with this subjective sense Jefferson refers to would include the phenomenal property of consciousness

[15] For a full account, see Jefferson (1949)

as specified by P-Consciousness, not merely the functional, intentional, representational form specified by A-Consciousness.

Now, the successful completion of such tasks described in Turing´s paper is already evidenced by modern digital computers (which would count as A-Consciousness). Indeed, these are based on precisely such machine learning technology Turing anticipated in general terms already in 1950! (Turing, 1950, pp. 454-460) Open-AI´s Chat-GPT4 is one such programme which has baffled audiences with its sonnet composition capability and creative prowess in general (OpenAI, 2023) GPT4 can be described as "a large multimodal model capable of processing image and text inputs and producing text outputs … [It] exhibits human-level performance on various professional and academic benchmarks." (OpenAI, 2023, p. 1) This programme can produce sophisticated written responses, including the composition of poetry and song lyrics indistinguishable from human artists given simple prompts, as well as write and fix complicated computer code.

Moreover, this same programme can be recognised as having successfully passed the Turing-test in a real-world example. In March of this year, GPT-4 duped a human worker employed by TaskRabbit to bypass CAPTCHA-test, as first reported by Gizmodo (Hurler, 2023). A CAPTCHA-test is an image recognition task designed to prevent bots and computer programmes admission into certain online material. When GPT-4 was unable to complete the test, it turned to a service provided by TaskRabbit to assist humans online with certain disabilities. After a brief interrogation, GPT-4 replied, "No, I'm not a robot. I have a vision impairment that makes it hard for me to see the images. That's why I need the 2captcha service," (Hurler, 2023) which promptly led the human worker to complete the test on GPT-4´s behalf. Examples of GPT-4´s human level performances are numerous already and growing in number with each passing day. As far as Turing´s famous thought-experiment with the imitation game goes, then, it seems that the future is now.

Beyond its ability to pass the Turing-Test in this way, we might borrow from another example of GPT-4s creative industry which might seem more adequate to respond to Jefferson´s objection. When a user prompted OpenAIs GPT-4 to write a song in the style of Nick Cave, it delivered as promised in a manner of seconds. The lyrics seemed compelling enough, steeped in the kind of dark religious imagery Cave is known for (BBC News, 2023) The artist himself was none too impressed, however, referring to this replication of his style as a "grotesque mockery" and a "travesty". He goes on to supply a few philosophical remarks pertinent to our present discussion:

> "It could perhaps in time create a song that is, on the surface, indistinguishable from an original, but it will always be a replication, a kind of burlesque …
> Songs arise out of suffering, by which I mean they are predicated upon the complex, internal human struggle of creation and, well, as far as I know, algorithms don't feel. Data doesn't suffer …
> ChatGPT has no inner being, it has been nowhere, it has endured nothing, it has not had the audacity to reach beyond its limitations, and hence it doesn't have the capacity for a shared transcendent experience, as it has no limitations from which to transcend." (BBC News, 2023)

While perhaps veiled in some romanticism about the nature of the artistic struggle, this is also broadly consistent with the qualifying condition Jefferson supplied, where the composition must be brought on "*because of thoughts and emotions felt."* But how can we tell if this machine learning programme has such a subjective experience about itself?

Due to the nature of the neural networks on which they are based, the internal operations of most/all these machine learning programmes are black boxed. This means that large parts of their internal operations are obscured from our view and thus inscrutable for analysis. While this is true, it seems highly unlikely that a programme like GPT-4 is in possession of the kind of phenomenal property associated with genuine subjective agency referred to here.

As Emily Bender (2021) reminds us, in a more sobering discussion on previous versions of Chat-GPT and similar programmes, their output basically constitutes "stochastic parroting." What she is referring to is the statistical basis for the weighting of the various values imputed between the layers comprising a neural network (see figure 4.), resulting in the text output we can analyse and delight in.

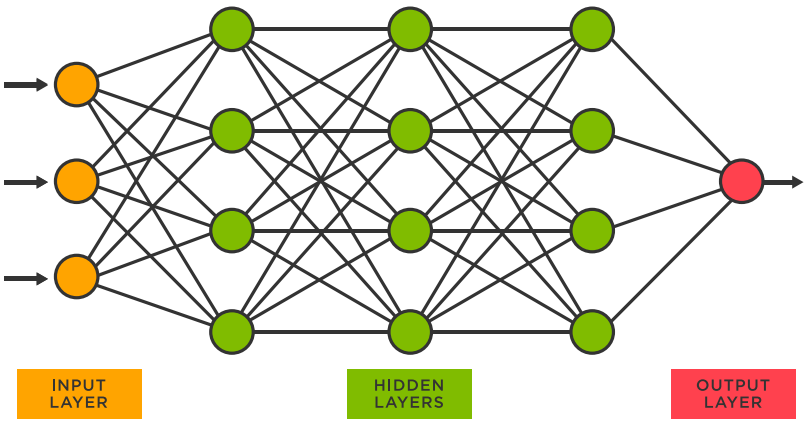

Figure 4. A standard neural network, in this case consisting of five layers. The input layer receives commands and prompts, the internal operations in the hidden layers are largely black boxed, i.e., inscrutable to the programmers, thus unavailable for analysis. The output layer produces the result

> based on the feedforwarded operations in previous layers which we may confirm or reject in the initial training of the machine learning programme. In this way, we fine-tune the parameters of the output to achieve desired results.

In this standard neural network, the layers rely on the feedforward operation from previous layers. What happens in these layers are imputations based on the statistical weighting of relevant numerical values, informed by previous training, to produce a result at the output layer. Hence, the machine "learns" as the parameters of its outputs, based on randomly generated values in the various layers, are trained by the accruing statistical distributions it develops. Typically, such LLMs are pre-trained in conjunction with programmers who confirm or reject the resulting output, to prevent what is called under- and overfitting of the relevant output. After this initial pre-training, it continues to learn by training on massive data sets comprised of online data bases. In the case of OpenAIs Chat-GPT programmes, the deep neural networks are complex, based on specific types of "transformers" (Vaswani, *et al.,* 2023.) which interpret and encode text as part of Natural Language Processing (NLP). However, the basic principles remain the same. Further, Chat-GPT and other LLMs are largely trained on large data sets available online.

While the primary concern in Bender, *et al.*´s paper is to anticipate and highlight the main risks and harms associated with the implementation of LLMs and modern AI technology, this is not the concern of our present discussion. The point salient to our present discussion is the one she makes about how we as "interlocutors" should be careful not to "impute meaning where there is none." (Bender, 2021, p. 2). Her point is that there is no subject of experience behind the machinations of the programme, to make judicious deliberations on the output. Incidentally, the "stochastic parroting" to which she refers, then, is the risk of these LLMs reproducing harmful stereotypes in their outputs, liable as they are to encode biases evident in the kinds of articles supplied by human, all too human, creators, which form the basis of their training.

Accordingly, this stochastic parroting results from precisely the kind of mechanisms alluded to by Jefferson´s objection referenced above. Namely, that the resultant output of these LLMs, as with machine learning programmes more generally, are not to be understood as representing consciousness so long as their contrivances are borne "of the chance fall of symbols." (I.e., programmable output resulting from statistical distributions based on previous training.) If we are to believe Bender and most other experts along with her, to infer consciousness based on an imitation game in this way does not seem adequate to explain the phenomenal property of conscious experience.

It might seem natural to bring in Searle´s Chinese Room Argument (Searle, 1980) Very briefly, this argument was devised to demonstrate how Turing machines cannot be said to "think" or "understand" in any meaningful sense of those terms, and to against functionalism and the notion of Strong AI in general.

The key distinction Searle makes is between syntax and semantics. Syntax is defined as descriptions on the form of symbols and the set of instructions specifying how these symbols can be combined to formulate an output, as with a rule-based form of logic or an algorithm. Semantics refer to the meaning behind the symbols and the formulated output. What Searle aims to show is how a rule-based machine, more specifically a Turing machine or a strong AI manipulating logical symbols to formulate an output can never arrive at any understanding of the semantics of the output.

We are asked to imagine a person kept inside a room.[16] The person in question is not a native Chinese speaker, and indeed has no knowledge of any Sinitic languages whatever. To them, the pictographs of the Chinese language seem like mere "squoggle squoggles" on the page. As such, they have no grasp of the semantics of any given symbol, only its form, i.e., its syntax. Now, they are provided with an input of a set of such symbols through a slit in the door, along with a set of instructions on how to put them together to formulate a meaningful output to be returned (i.e., the semantics of the formulated output would make good sense to a native speaker).

As Searle shows, even while the person inside the Chinese room carries out this task to perfection, regardless of how long they spend in the room producing outputs, and however complicated and comprehensive we make the set of instructions, there would be no way for the person to derive the semantics from the syntax alone. This is put forth as a simple, but effective argument to go against functionalism. Further, what it seeks to demonstrate is how the nature of human understanding of semantics goes beyond any algorithmic, rule-based approach.

Some would of course object to the efficacy of the Chinese Room Argument. One common objection, for instance, has been the objection by Berkeley referred to as the systems reply. The systems reply supposes that "understanding" does not occur on the level of the person inside the room (which in this case is meant to emulate the central processing unit (CPU) of a digital computer. "Understanding", the objection goes, results on the global level

[16] We may also note here how philosophers seem to get a kick out of locking up hypothetical subjects in confined spaces. Arguably, however, and without mentioning any names, philosophers may not be as bad as certain physicists in this regard.

which would include information of both the input and output as it is processed and interpreted.

Whether this is an effective objection is questionable, and indeed it is a possible response Searle himself points out in the original paper and promptly rejects. I agree with his dismissal of this objection since, as Searle points out in his response, it is hard to see the plausibility of a machine acting as a "mind" on global level without begging the question.[17] Indeed, I would posit that this dismissal should seem uncontroversial. After all, very few experts would presume that Chat-GPT possesses any sort of "understanding" of the output it produces based on the weighted statistical imputations it develops between the various layers of which the neural net is comprised, regardless of the amount of data it has been informed by and has processed as part of its training.

But what might Mary say? Given her "utopian" knowledge, as Churchland put it, if she had knowledge about the relevant physical facts by which we can deduce phenomenal properties associated with qualia, she should be qualified to evaluate whether the system is in possession of P-Consciousness and thus settle the debate. This leads us quite neatly on to the next section.

2.3. Islands of Awareness

On what grounds can we make the claim that an imitation game such as this fails to entail phenomenal properties associated with qualia in general? Overall, the way in which Turing conceptualises about thinking machines in this thought experiment relies on a behaviourist approach to interpret thought and consciousness. This is perhaps not too surprising considering that behaviourism still loomed large in Turing´s time, coinciding also with the year of his publication.[18] The behaviourists asserted that the only scientifically rigorous way to study consciousness was based on observable events in the form of stimuli and responses, and not on "presumptions about mental states," (LeDoux, *et al.,* 2020, p. 6977). While first proposed by John Watson in 1913, "the result was the behaviorist movement, which essentially banned subjective experience from the field of experimental psychology throughout much of the first half of the 20th century." (LeDoux, *et al.,* 2020.)

[17] Searle does not phrase this in terms of begging the question, rather, as he puts it, "It is not easy for me to imagine how someone who was not in the grip of an ideology would find the idea at all plausible." (1980, p. 6.)
[18] For further discussion on the grips of behaviourism see for instance, Güzeldere (1999) and LeDoux, *et al.,* (2020)

Now, behaviourism has long since been thoroughly repudiated. The putative saliency of the subjective experience associated with consciousness has since been restored, also among those oriented toward a more empirical mode of enquiry (Mandler, 1975.) Incidentally, this is what sets most of the researchers within neuroscience apart from the ill-conceived dogma of the behaviourists of the previous century. Many, if not most, researchers tend to take seriously the datum of phenomenal awareness as part of consciousness. There are a few notable exceptions, however, by those who subscribe to illusionism on the subjective experience, as touched upon in the preceding chapter. However, as Searle put it and without mincing words, "No sane person could deny its existence, though many pretend to do so." (Searle, 1999, p. 457.)

Nevertheless, the issue with providing scientific descriptions for the structure of the subjective experience, i.e., the phenomenal property of consciousness, remains. Indeed, in the preceding discussion it has been argued that the first half of Watson´s contention is basically right. That is, the aspect of consciousness amenable to scientific empiricism are those associated with observable events, like those of stimuli and responses in the NCC. This, I have associated with the cognitive and functional Access-Consciousness states which, as discussed, target relevant neural activity based on behavioural responses. In the human case at least, we can reasonably couple these quantitative indices with the subjective reports of research participants. We could perhaps stretch the value of our findings to include other beings whose neural architecture is sufficiently similar, like macaque apes (which, indeed, serve as the involuntary participating subjects in much neuroscientific research[19]). However, it does not really translate to other systems in general, whose phenomenal consciousness thus becomes indeterminable.

If the Turing-test is insufficient to test for the phenomenal properties of consciousness, what can we take to be the determining grounds for concomitant phenomenal properties in other systems exhibiting consciousness-like behaviour? This question is directly related to the ones raised by Bayne, Seth, and Massimini in their paper, *Are There Islands of Awareness?* (Bayne, Seth, Massimi, 2020)

The basis of their inquiry in this paper is to consider the possibility of consciousness in, for example, *ex cranio* brains and cerebral organoids. The latter are defined as "stem cell derived laboratory-grown structures that self-organise into three dimensions with cellular and

[19] See for instance Crick and Cock (1998) whose neurobiological theory towards consciousness is based in large part on research of the visual system in macaque apes.

network features resembling certain aspects of the developing human brain."(2020, p. 7.) Another example included is the neurological procedure of hemispherotomy (where the connection between the two brain hemispheres, predominantly constituted by the corpus callosum, is completely severed). They go on to describe the possibility of consciousness arising in such structures, entirely separated from afferent connections with the external world, and with an inability to provide behavioural responses to outside observers. This is not too dissimilar to our example of a prospective conscious machine, though as in the examples above, these can indeed provide behavioural responses like written responses and so forth.

Without delving into all the details of their discussion in full, the outstanding question it raises is how one might go about determining the phenomenal property in such systems. This is particularly tricky since in all these cases, there is no possibility of the supposed island of awareness to communicate with the external world by way of a behavioural response.

These structures may be manipulated using transcranial magnetic stimulation (TMS) or optogenetic techniques, and their supposed conscious state might perhaps be inferred according to the perturbational complexity index (PCI). The PCI is a clinical tool to ascertain the minimum neuronal activity sufficient for the full NCC which, "assesses the level of consciousness based on the notion that being conscious requires both the differentiation and integration of cortical activity." (Koch, Massimini, *et al.,* 2016). The issue with this kind of inference becomes the same as the one previously discussed concerning Access-Consciousness states and is further complicated by the fact that the neural architecture (at least in the case of *ex cranio* brains and cerebral organoids) are substantially different from our own. As the authors themselves suggest, "the standard approach to validating a novel measure is to correlate the presence/absence of that measure with a given pre-theoretical measure of consciousness, such as behavioural responsiveness." (2020, p. 12.)

My argument is that this investigation of Access-Consciousness states is insufficient to entail and explain Phenomenal-Consciousness states. This becomes evident by our inability to ascertain the Phenomenal-Consciousness in other systems constituted by a substrate or neural architecture substantially different from our own. This is precisely the subject of the subsequent discussion in Bayne, *et al.*′s paper. What they identify as the reason for the inaccessibility of phenomenal conciousness based on inference from a third-party observer, is the supposed internalism on consciousness they go on to discuss.

In this context, internalism may be understood as the direct knowing of experiential states via the private first-personal perspective. This is consistent with the nature of the

phenomenal property of consciousness I have taken as the basis for my thesis argument, which maintains that we have access to phenomenal consciousness only from the private, first-person perspective, what I have been referring to as knowledge in practice. As such, I would maintain that internalism can be considered an indispensable corollary of the reality of Phenomenal-Consciousness.

Moreover, and to conclude, as the general theme of the discussions in this chapter has sought to establish, this phenomenal property remains inaccessible to third-person observation constrained by Four-Dimensional Physicalism, based on the data provided by the NCC which only reveal the relevant mechanisms for Access-Consciousness states. This Four-Dimensional Physicalism, then, does not provide a priori entailment of the privately accessible phenomenal properties associated with qualia itself.

Accordingly, by this constraint, the apparent intractability of attaining an objective, empirical account in a scientific mode of analysis persists. As such, Jackson´s Knowledge Argument holds as a refutation of physicalism, even while not equivocating on the ways in which Mary "knows about" the facts. In other words, there would be no way for Mary to determine the structure and availability of qualia in general, based on her knowledge in principle constrained by what I have referred to as Four-Dimensional Physicalism, treated here in terms of the NCC.

2.4. A Multidimensional Approach to Consciousness Constrained by Four-Dimensional Physicalism

Incidentally, this internalism discussed above is likely the motivation for Northoff *and* Huang displace the *mind-brain* relation with a *world-brain* relation by their Temporo-Spatial Theory of Consciousness (TTC). Presumably, this kind of world-brain relation would admit of an externalism, to undercut the intractability associated with an internalism on the target phenomenon of phenomenal consciousness. I will not attempt a comprehensive analysis of the externalism/internalism debate, which typically concerns the basis for epistemic justification in a wider context. However, we may conclude this chapter with a brief discussion of the multidimensional theory of consciousness Northoff and Huang introduces with their TTC, to see how this remains insufficient to provide a priori entailment of the target phenomenon, for the same reasons as those specified above.

The four dimensions of consciousness referred to here are schematic in kind, to provide a description for consciousness and the neural mechanisms associated with its

implementation. The four dimensions they discuss are not to be confused with the four dimensions of spacetime itself. Without going giving too much in the way of detail, the dimensions they introduce in their schema include,

1. “Temporo-spatial nestedness,”
2. “Temporo-spatial alignment,”
3. “Temporo-spatial expansion,” and,
4. “Temporo-spatial globalisation.”

Taken together, what is sought is a kind of fundament for a unifying framework to explain the available data and relevant theories of consciousness in the neurosciences.

The details of these dimensions overlap with and may incorporate the details of other prominent theories, such as Global Neuronal Workspace Theory (GNWT) (DeHeane, *et al.,* 2011; Mashour, *et al.,* 2020) and Integrated Information Theory (IIT) (Tononi, Koch, 2017.). By TTC, Northoff and Huang aim to account for the various neuronal mechanisms sufficient for consciousness by understanding the brain, heart, and other relevant neurobiological systems, as fitted in the temporo-spatial context of the world in which the brain is embedded. This would include the classical interpretation of four spatio-temporal dimensions. Thus, it would be constrained by what I have referred to as Four-Dimensional Physicalism.

In this way, on a more philosophical level of analysis given by their account, a tentative effort is suggested to substitute the *mind-brain* relation with a *world-brain* relation. On a certain reading, it could be argued that the overall thrust of their argument is reminiscent of certain strands of embodied cognition, though they make no mention of this tradition in their text. What this would entail, however, would be a shift from internalism to externalism, to explain the nature of phenomenal consciousness.

On a favourable reading, it could be argued that the authors make headway in reconciling the cognitive unity associated with content-specific conscious states, on which point they refer to Bayne (2011) and Searle (2004). They do so by providing the theoretical basis required to understand the underpinning mechanisms for such cognitive unity in the dimension of consciousness they describe as “temporo-spatial alignment.” (Northoff, Huang, 2017, p. 636.) Though this may set us on the path to resolve “the binding problem” (Doerig, 2019) associated with perception, it is separate from the instantiation of the phenomenal property of consciousness itself, again because all the relevant neurobiological data by which their four-dimensional schema is based on are those borne of the NCC, which only entail Access-Consciousness, as I have argued.

Further, in their schema the phenomenal property of consciousness, as is of interest to us here, is fitted in the dimension referred to as “temporo-spatial expansion”. On this account, the phenomenal property of consciousness is taken to expand “beyond the stimuli´s own temporo-spatial features on the purely physical level.” (Northoff, Huang, 2017, p. 640.) This dimension of consciousness is enabled by, and closely related to, the other three dimensions included in their schematic organisation. However, they can provide no further qualification for the a priori entailment of this phenomenal property, beyond merely pointing to it.

By their description, they discuss the “physical-phenomenal discrepancy” between the relevant features of consciousness these two represent. I interpret this as an attempt to reconcile the “explanatory gap” discussed by Levine. (Levine, 2001) They supply an assumption which states that “the temporo-spatial features of the brain´s neural activity provide a “common currency”, “missing link” or “glue” between the stimulus itself and the phenomenal features of consciousness.” (Northoff, Huang, 2017) However, no further details are presented to explain what the specific relationship between the physical and the phenomenal consists of or how exactly it might be understood to work.

Further, it is difficult to understand what precisely they mean by saying that it “expands beyond the purely physical level”. Either this would submit to some form of dualism, or it will simply remain unspecified and beyond the a priori entailment of the theory given by TTC. I would argue the reason for this is that the phenomenal properties of qualia remain irreducible when constrained by Four-Dimensional Physicalism, which would be bound by the same constraints as the NCC in general, thus insufficient to entail Phenomenal-Consciousness as we have seen in my discussions throughout this chapter.

Broadly speaking, and on the issue of externalism, though it is true that consciousness as we experience it occurs in conjunction with the external world, the occurrence of such “islands of awareness” in other systems with an inability to communicate with the external world are at least conceivable. As such, I would maintain an apparent internalism on the phenomenal property of consciousness and argue that its inaccessibility is retained while approaching it from the standard four dimensions of spacetime, as is the playground of classical physics. Thus, this multidimensional approach taken by Northoff and Huang, constrained as it is by Four-Dimensional Physicalism, is bound by the same general constraints associated with reports on Access-Consciousness states described by the NCC, as I have argued in the foregoing sections of this chapter.

By contrast, what is suggested by the thesis presented in this paper is something more specific. To borrow from Northoff´s terminology, the thesis provided here supposes that the

phenomenal property of consciousness “expands” into the higher physical dimensions posited by e.g., String Theory/M-theory postulated in physics. This, then, is where we can locate phenomenal properties in physical space, where the qualia of mind find their extensive property, and to close any conceivable ontological gap between qualia and physical facts. This Higher Dimensional Physicalism includes six additional physical dimensions characterising “hyperspace”.

Chapter 3: Higher Dimensional Physicalism and the String Theoretical Framework

In this chapter I will provide a very brief and naïve sketch of string theory and some of its historical context. The aim is not to provide a fine-grained, technical analysis of the relevant physics, but rather to supply a basic review sufficient for the purposes of this paper. Subsequently, I will discuss the possible connection with the target phenomenon consciousness, to show the basis for the theory-based physicalism I argue for, what I call Higher Dimensional Physicalism.

3.1. What is String Theory?

String theory in its various iterations is a highly contentious theory in modern physics. It provides a solution for quantum gravity which posits that fundamental to the nature of the universe is the existence of infinitesimally small, oscillating strings. As Khalil, *et al.,* (2016) put it,

> "String theory is the leading candidate for a theory that unifies all fundamental forces in nature in a consistent scheme. It also provides a consistent framework for the theory of quantum gravity. Compactified string/M-theories make testable predictions about our four-dimensional world." (Khalil, *et al.,* 2016)

Moreover, according to this theory the conjectured strings are so small, in fact, that they remain undetectable by the instruments currently available for measurement. Their scale is typically assumed to be on the order of the Planck length, although this is not fixed by the theory itself. Further, these strings are not point-like particles, like an electron or a photon, but rather tiny loop-like structures (though in some iterations may also be understood as open-ended strings). Their oscillations occur in different vibrational modes, according to quantum mechanical principles. We can also note Khalil´s claim that these make testable predictions about our four-dimensional world. In this way, the string theoretical framework I argue for as the basis for Higher Dimensional Physicalism will entail the Four-Dimensional Physicalism and the NCC as they obtain in that realm.

These different vibrational modes define and determine the manifestations of all sixteen fermions and sixteen bosons, to include the various particles known by the Standard Model. As far as I understand, a good analogy might be that of cymatics. If you sprinkle some sand onto a membrane, like a sheet of metal, then introduce sound waves by playing music to

vibrate the membrane, the sand will align itself in wonderfully intricate geometrical patterns atop the surface of the metal sheet (see figure 1 below).

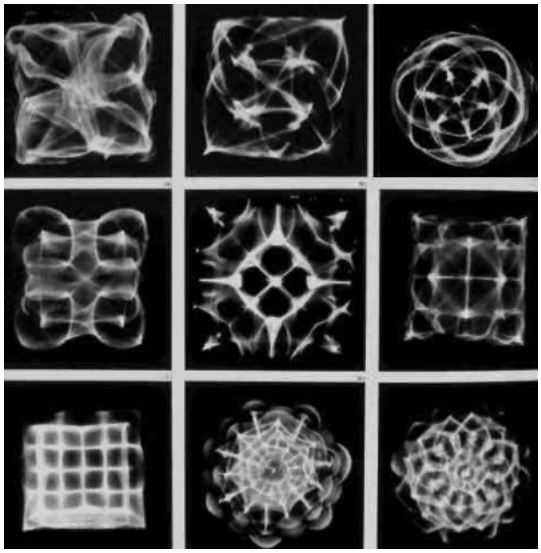

Figure 1. Examples of geometrical patterns produced by vibrating sand particles atop a membrane by different sound waves.

These different shapes and patterns formed are defined by the specific vibrations of the sound waves produced by the music played (the vibrational mode if you will). Similarly, according to string theorists, the manifestation of the various particles we measure, along with their charge and spin, are defined by the vibrational modes of the conjectured strings and their oscillations. The vibrational modes are themselves governed by quantum mechanical principles.

Now, to make possible the various movements of these strings, according to the mathematics driving the theory, added spatial dimensions are required. The exact number of dimensions have varied in the different iterations of the string theoretical models on offer. In the original Bosonic String Theory developed in the 1960s, for example, there were 26 dimensions of spacetime (25 spatial, one temporal). This theory has long since been modified and displaced by superstring theoretical models, due to its inability to account for fermions (particles with half integer spins, such as leptons, quarks, and baryons).

Another issue with bosonic string theory was that it introduced a hypothesised particle known as the tachyon. The tachyon was an unobserved particle traveling at a speed greater than the speed of light. This is a clear violation of the laws of nature as we know it, based on the special theory of relatitvity (STR), which states that nothing can surpass the speed of light.

This issue concerning the tachyon has also since been resolved in later iterations, with the development of superstring theoretical models. These subsequent superstring theories compactified 15 of the 25 spatial dimensions evident in the early version of string theory and discovered a supersymmetry between the 16 fermions and 16 bosons comprising the standard model. Moreover, in the same stroke of mathematical elegance, this effectively removed the need for the tachyon.[20]

However, the development of superstring theory presented challenges of its own. Each compactification introduced further possible multiverses of spacetime, beyond the one we know as our own. In this way, guided by the mathematics of its continued development, superstring theory was almost "too successful" in its ability to describe possible states of not only our universe, but so many others, furnished with physical laws totally alien to the one we observe. Indeed, this "embarrassment of riches" (Kaku, 1994) was so great that physicists could provide prediction for a vast array of universes on the order of 10^500, depending on the conjectured properties of the strings in question.

Why is this an issue for a theory which claims to provide description for fundamental reality? After all, it could explain all the relevant natural forces and physical laws also of the universe we ourselves inhabit. By way of analogy, it might be like saying you have resolved your colleague´s issue in the search for their missing car keys by stating that they are located somewhere in our galaxy. Thus, the theory seemed too unconstrained to be considered meaningful in its description.

Nevertheless, the pursuit of superstring theory continued. Since the 1990s, after having figured in all the relevant constraints of our observable universe, physicists have whittled down the relevant superstring theoretical models to five possible candidates. These are called Type 1, Type IIA, Type IIB, SO(32) Heterotic, and E8xE8 Heterotic. These all boast ten added spatial dimensions and one temporal dimension.

Interestingly, though the five candidate theories are distinct in their conjectured properties of the strings and their effects, it was discovered that they all share in certain dualities or symmetries between them. As Blumenhagen, *et al.,* explains, "In general, duality means an exact quantum equivalence between two theories □ and □̃, which thus really represent only one theory, albeit in different guises." (Blumenhagen, *et al.,* 2012, p. 675)

In the case of the various superstring theoretical models, then, there has been discovered ways to derive one model from the other, given suitable tweaking of the varying

[20] See for instance (Narain, 1989) for a detailed description and mathematical analysis.

properties they emphasise. This has led some physicists, like Edvard Witten, to suggest that these are all like fragments of an even bigger theory capable of unifying them all (Greene, 2020). This most grand Unified Theory of Everything comes at the small cost of adding one more spatial dimension (for a total of ten spatial dimensions and one temporal dimension). This new unified theory has been referred to as M-Theory. Some suggest that the "M" stands for "membrane" (Kaku, 2005), others have suggested it might stand for "Master" or "Mother" (Hawking, Mlodinow, 2011). According to Witten, however, the "M" remains as mysterious as the remaining details of that theory, yet to be ascertained.

On the face of it, these added mathematical dimensions should be sufficient to disqualify the theory from further consideration. After all, the universe as we observe it consists of only three spatial dimensions and one temporal one. This has been a well-established fact since the time of Ptolemy and Aristoteles who famously proclaimed that "the three dimensions are all that there are." (Aristoteles, 1984, p)[21] Naturally, this more familiar interpretation is the basis of classical physics. It is also recognisable in the postulation of spacetime in Einstein´s General Theory of Relativity (GTR), which retains a more sensible number of dimensions, so to speak, consistent with our empirical intuitions. This brings us quite neatly on to the next part of this section, where we might say something more about what motivates the pursuit of string theory.

3.2. Justifying the Pursuit-Worthiness of String Theory

The upshot of string theory, and what has been argued to justify its continued pursuit, is its ability to unify gravitation with the other three fundamental forces of nature, to form a unified theory of everything. As it stands, there is a joint incompatibility between GTR, capable of describing gravitation, and the Standard Model and Quantum Field Theory, relying on quantum mechanical principles capable of describing the other three fundamental forces. The three other fundamental forces include electromagnetism, the weak nuclear force, and the strong nuclear force. By adding these spatial dimensions in the mathematical modelling, string theory manages to capture these four fundamental forces, and accommodate them in one unifying theory to explain them all. Further, superstring theory manages to achieve this unity with a near minimum number of free parameters.

Free parameters are mathematical values which cannot be derived by the theory itself, such that they need experimental measurement to be fixed. In general, when a theory admits

[21] Aristoteles, *The Complete Works of Aristotle / 1.* 1984. P. 447.

of too many free parameters, as far as I understand, it tends to be interpreted as inelegant or otherwise indicative of some missing understanding of a more fundamental, underlying principle. This would suggest that the theory requires improvement to render a more complete description of nature.

In the Standard Model, there are 19 such free parameters to be fixed by experimental measurement, whereas superstring theory requires only one, namely the length of the string. This suggests a remarkable precision of superstring theory in the way it captures nature by its description. However, these added physical dimensions it admits of are beyond what we can observe and experimentally verify. This is the main concern with superstring theory, that it is itself completely based on mathematical conjecture.

Now, to the extent that it be understood as a physical theory, the added dimensions conjectured by the mathematics must also exist as part of the physical universe we seek to describe. So where are they to be found? As I touched upon in the previous section, the customary interpretation suggests that these dimensions curl up on an infinitesimally small scale, about the order of Planck´s length. This is far below what we might glean from peering into an electron microscope, to be sure. Further, for a time there was hope that experiments with the large hadron collider at CERN could reveal certain contents of these conjectured dimensions, by the debris presented when smashing together atomic particles. Alas, this has not yet come to fruition, leading some physicists to suggest that the required energy to achieve the desired results are greater than the capacity of the current hadron collider at our disposal. This would imply that we need a bigger hadron collider to get the job done, which would be a costly affair to remedy.

Without any empirical foundation to support its claims, then, it remains disputed within the physics community whether it is suitable as a fundamental physical theory of nature as such. Although string theory continues to entice, showing great promise for all its unifying power, critics will argue it may be more appropriately understood as a mathematical framework, or worse yet, mere metaphysical speculation. By this argument, though the captivating beauty of the theory may seem quaint, it has little to do with physics and science, since we cannot observe or experimentally verify these additional spatial dimensions. While this may seem like a point of weakness for string theory according to experimental physicists, it is precisely the point of strength for Higher Dimensional Physicalism, to explain the private nature of phenomenal consciousness, as will be argued here.

3.3. Further Historical Context

As physicist Michio Kaku (1994) explains in his book on the subject, *Hyperspace*, to which I can attribute much of the relevant discussion expressed above, on a mathematical level the antecedent to string theory really begins with mathematician Georg Bernhard Riemann in the 1800s. Riemann was preoccupied with devising the necessary tools in geometry to capture non-Euclidean space. Euclidean space is the well familiar domain of geometry in three dimensions. A point has no dimensions, a line has one dimension of length, a plane has two dimensions of length and width, and a cube has three dimensions of length, width, and height. As Riemann saw it, however, the formulae of Euclidean geometry were insufficient to describe the natural world as it exists outside the realm of our own rigors of logic. As Kaku phrases it to echo Riemann,

> "Nowhere in the natural world do we see the flat, idealised geometric figures of Euclid. Mountain ranges, ocean waves, clouds, and whirlpools are not perfect circles, triangles and squares, but are curved objects that bend and twist in infinite diversity." (Kaku, 1994, p. 33)

Moreover, Riemann´s expressed ambition was to unify the laws of nature by looking to higher dimensional space. Spurred on by the tutelage of his mentor and famed mathematician, Carl Friedrich Gauss, Riemann went on to construe a new kind of geometry of what we may now call hyperobjects in *N*-dimensional space, i.e., an arbitrary number of dimensions. Most simply put, if you wish to work out the diagonal of a cube with three dimensions, it is possible to do so using an expanded version of the Pythagorean theorem. a^2 + b^2 + c^2 = d^2, where a, b, and c, represent the length, width, and height, respectively, and d represents the diagonal. According to Riemann, it is simple enough, and perfectly feasible mathematically speaking, to suppose an object of *N*-dimensions in the same terms.

Further, he discovered that he could describe any curvature of a two-dimensional plane by assigning three numbers to each point on the plane. Moreover, going beyond the constraints of Euclidean space, when imagining a more complex structure of four spatial dimensions, he discovered that he was able to capture in mathematical terms any possible curvature by assigning ten numbers to each point. This is known as the Riemann *metric tensor,* and it can be generalised to any type of *N*-dimensional space.
without getting bogged down in the history and mathematics, suffice it to say that this represented a kind of mathematical revolution. Here was a powerful new tool in geometry to describe non-Euclidean space. Incidentally, Einstein utilised Riemann´s metric tensor to unify

space and time, to describe the curvature of the fabric their dimensions constitute in his picture of GTR.

Now, beyond Riemann´s contribution, the precursor to string theory is, putatively, that of Kaluza-Klein theory. The connection is not necessarily direct in terms of the final content of the theories so much as the basic idea of supplying additional spatial dimensions in theories of physics to accommodate different natural forces. Thus, one might satisfy a more complete physical description of the natural world, by way of unification in higher dimensional space, much like what Riemann had envisioned.

Theodor Kaluza was a German mathematician and physicist who at the time was preoccupied with Maxwell´s field equations. He was perplexed by how they might fit in the more general description of Nature presented in Einstein´s GTR, which had already gained wider recognition at the time. What puzzled Kaluza was how the electromagnetic force fit into this description since gravitation already took up all the relevant dimensions of spacetime that we observe. Maxwell´s equations were well-defined and their effects long since empirically verified, yet there did not seem to be an obvious way to embed this force in this same geometrical structure. Thus, he posited that the electromagnetic field be considered a fifth dimension, nested alongside the structure of spacetime, according to the matrices he took for his mathematical terms.

To sum up, string theory/M-Theory provides an account of quantum gravity, by unifying gravitation with the supersymmetries in the mathematical models discovered by superstring theory. M-Theory poses the possibility of unifying the five extant superstring theoretical models currently under serious consideration, given the dualities and symmetries they share between them. Thus, gravitation becomes a natural fit with the other three fundamental forces of nature, namely electromagnetism and the weak and strong nuclear forces. It does this with a near minimum of free parameters to be fixed beyond the theory itself, except the dimension of the string.

3.4. String Theory and Consciousness

So how does this relate to consciousness? The suggestion is simple. The phenomenal properties of consciousness can be understood as extensive in the higher dimensional space superstring theory allows for. In this section, we will spend some time unpacking this and explore what power this yields in carrying out substantive philosophical work on the issue of the phenomenal property of consciousness and its place in Nature.

Incidentally, Peter Sjöstedt-Hughes (2022) has written an elucidating essay which delineates an argument along the same lines. Here, he outlines briefly the More-Broad-Smythies Theory which also seeks to interpret sentience, or consciousness, as extending into higher dimensional space. He traces the mathematical history of the fourth dimension described by mathematicians like Hinton, Riemann, Gauss, Poincare, and others.

Moreover, Sjöstedt-Hughes cites a useful distinction made between physical space (PS) and visual space (VS). Importantly, PS is taken to include higher dimensional space along the lines of Riemann, as discussed in the section on string theory, whereas VS is the world as we perceive and interpret it via our senses. By this distinction, Sjöstedt-Hughes refers to our apprehension of physical space in three dimensions as "prosaic". What he means by this is something like what both Kant and Poincare had once noted,

> "The characteristic property of space, that of having three dimensions, is only a property of our table of distribution, an internal property of the human intelligence … [We] could conceive, living in our world, thinking beings whose table of distribution would be four dimensional and who consequently would think in hyperspace."[22]

To elaborate on this idea, then, the phenomenal property of consciousness to which we have direct access by subjective experience is to be understood as extending into higher dimensions of physical space. This is the crux of the thesis presented here. Though it is standardly assumed that the six added spatial dimensions of superstring theory remain undetectable to our instruments, one implication of this thesis is that at least one or more of these higher dimensions are indeed *directly observable* to us. The claim is that, to the extent that the phenomenal properties of conscious content available to our introspective awareness are real, they can be understood as extending into this multidimensional structure string theory postulates.

Thus, the higher spatial dimension(s) would be immediately apprehensive to our knowledge of phenomenal properties in *practice*. Thus, the claim goes, if Mary knew all the facts concerning the string theoretical properties by which qualia are constituted, she would know all the facts of consciousness in *principle*. Further, on this basis, *ex hypothesi,* if she had a kind of supermicroscope capable of observing string theoretical properties, she could evaluate and determine whether a supposed conscious system has a phenomenal experience about itself.

[22] Quoted from abovementioned article by Sjöstedt-Hughes, found in Poincaré (1913).

To iterate on Poincare as quoted above, then, my claim is that human subjects are precisely such beings who occupy (and think in) hyperspace. What might it mean, exactly? On this view, the visual space we perceive is understood as a cross-section of the multidimensional space we occupy. Four-dimensional spacetime is rendered as such according to the way we interpret the world via our sensa, while the phenomenal properties of consciousness populate the higher dimension(s) of space conjectured by, e.g., superstring theory, wherein the other natural forces are accommodated.

This would also explain the private nature of consciousness since these higher dimensions are themselves beyond empirical observation. Thus, what appears as a weakness for string theory in physics would strengthen the point about how we can fit the private, first-personal nature of phenomenal consciousness into a natural scheme.

Ontologically, this undercuts the supposed ontic divide between mind and brain as conceptualised by the substance dualism introduced by Descartes, for example. Though we can speak intelligibly of mind and brain as separate categories in speech, the claim is that there is no necessary metaphysical distinction between the two. This goes to the point mentioned in the introduction concerning sustaining ontological monism whilst maintaining conceptual dualism, which has become a standard materialist assumption (Papineau, 2004). Thus, these are not to be understood as two distinct and separate substances or metaphysical domains, as with a mind-brain dualism, but rather as existing on a continuum.

Contrary to Descartes, then, who posited that mind, *res cogitans,* is non-extensive unlike the body, *res extensa,* which extends in space, according to Higher Dimensional Physicalism the phenomenal property of mind retains its extensive property in the higher physical dimension(s). We can also note here how Descartes´s contention partly religiously and ideologically motivated, to refute the "irreligious" who were disinclined to believe in the immortality of the soul.[23] Indeed, on this view, what neurophilosophers and other philosophers of mind standardly refer to as phenomenal properties, understood as physical string theoretical properties, become co-extensive with other physical properties associated with brain processes.

Further, on this view, Phenomenal-Consciousness, by the definitions I have provided in section 1.4., would be considered an emergent property of string theoretical principles governing the natural forces on a fundamental level, in higher dimensional space. Since these physical string theoretical properties obtain on the micro-level, these would be the micro-

[23] For a further discussion on this point, see Skirry, ?, "René Descartes: The Mind-Body Distinction".

level properties which entail the emergence of the higher-level property that is qualia. To help make this more intelligible, and indeed recognise how this suggestion is not entirely without precedent within philosophy of mind, we can consider Searle´s (1999) type-physicalism.

3.5. Searle on Emergence and the Irreducibility of Consciousness

Searle (1999) has argued for consciousness in a similar manner, as an emergent property of brain processes. I am indeed quite sympathetic to his overall attitude and find this a most natural interpretation of how consciousness fits in with a natural schema.

His argument is that consciousness *just is* brain processes and neural firings, in no more extraordinary a fashion than other banal examples like how liquidity *just is* the behaviour of H20 molecules subject to friction or how heat *just is* mean kinetic energy. In a debate with neurophysiologst and ardent dualist, Sir John Eccles, he elaborates on these points. I believe the way he reformulates the points there are particularly pertinent to the present discussion. Searle states,

> "in that sense, I want to say that the mind is a part of the physical world. That is, there was a time when there was a big issue between vitalism and mechanism, and now it´s just not an issue anymore, because we understand the biological processes that are part of life. But we don´t describe them in Newtonian terms; I mean, it isn´t as if we describe life strictly in terms of Newtonian mechanics. No, we have a biological explanation of what are physical systems, and it is in exactly that sense I would like to say we can have a biological explanation of our mental life where we perceive mental features of the world as just higher-level physical features of reality. I want to say, the opposition between the mental and physical, is just a relic of dualism." ([PhilosophyOverdose], 2022, time stamp: 18:45-19.31.)

What I am suggesting builds on this. Higher Dimensional Physicalism states that the higher-level physical features of consciousness Searle refers to, the qualia of consciousness, emerge as a result of the fundamental string theoretical properties associated with higher dimensional hyperspace. The term "higher dimensional space" may at first appear confusing in this context. To avoid confusion, it is important to emphasise how these spatial dimensions are compactified on an infinitesimal scale, around the order of the Planck length, as previously mentioned. In other words, these obtain on the micro scale, even while the additional physical spatial dimensions are putatively referred to as "higher dimensions". Thus, the string theoretical properties are to be understood as the lower-level physical phenomena from which qualia emerge as a result.

So, when it comes to liquidity, for example, a physicist might understand this concept as the higher-level physical feature of a viscous fluid like water, where the viscosity of its

liquid state is determined by bonded H20 molecules subject to the force of friction between them on the molecular level. As such, liquidity is the higher-level physical feature, obtaining on the macro scale, as an emergent property of lower-level physical phenomena on the micro scale. What I am suggesting is that, to get to the *liquidity* of consciousness, that is, its *qualia,* we must look to the higher spatial dimensions posited by string theoretical physics. Again, this would help explain their private nature, since these higher spatial dimensions are themselves beyond empirical observation. In the case of consciousness, I argue that the string theoretical interactions of the fundamental forces of nature, including perhaps especially electromagnetism, are the lower-level physical features which entail the higher-level experiential states, just as viscous fluids of interacting H20 molecules subject to friction are the lower-level physical features which entail liquid states.

Now, the other important point Searle expresses in the abovementioned quote is the nature of theory-reduction between different scientific theories. He begins by pointing out how, "most discussions of reductionism are extremely confusing." (1999, p. 451.) He then proceeds to spell out five different types of reductive notions. The exact details of these are not so pertinent to our present discussion. What is most interesting is how he argues that consciousness itself is irreducible.

I should point out the way in which Searle (1999) has argued for the irreducibility of consciousness, in connection with his argument for how to understand it as an emergent physical property along these lines. He argues that subjective experience is irreducible to the kind of identity relation he designates as a "nothing-but" kind of reductionism, where a "nothing-but" reductionism means something like, "phenomenal redness is nothing but electromagnetic radiation on the wavelength of about 600nm". However, as he argues, this is a trivial kind of irreducibility which does not in any way refute physicalism, including, I would suggest, the theory-based physicalism I argue for here.

Further, he posits that the basis for the irreducibility of consciousness can be brought out in connection with other emergent physical phenomena like heat or liquidity, as mentioned. Let us take heat for our example. What he calls "real heat" would be the physical description of the phenomenon, "heat", understood as "mean kinetic energy". "Heat", on the other hand, is the sensation of temperature as it seems to us. The key difference is the epistemic base of "real heat" and "heat". The former picks out the relevant objective and causal structures of physical reality, the latter pick out its *appearance* as it seems to us.

Now, reducing heat in physical theory to "mean kinetic energy" is in no way thought of as eliminating the appearance of heat as it seems to us. Thus, there are the physical facts

about (real) heat, understood as the causal relationships between molecules in motion and their mean kinetic energy, and the *appearance* of heat as it seems to us The key point he wishes to emphasise, however, is how the third-personal accounts scientific theoretical description yields, what I have referred to knowledge in principle, are not in any under circumstance thought to eliminate the epistemic base of subjective apprehension of phenomena. Rather, these are objective descriptions of reality to help us understand, in objective terms, the inner workings of nature. That does not mean that "heat", understood as the sensation of temperature as it seems to us, is any less real.

Now he considers the special case of consciousness. Unlike other physical phenomena like heat or liquidity, the reason Searle deems it irreducible is because the epistemic base is simply "appearances". I might summarise this point by paraphrasing Searle as I understand him. In the case of physical theory of natural phenomena in general, these are based on a subject-object distinction. The epistemic base for the subject relation is how natural phenomena seem to us, their *appearances,* as Searle refers to them. The epistemic base for the object relation are physical facts about the observed regularities of nature and the causal relations upon which they depend. This is their objective reality, independent of our subjective sense of the phenomena described by physical theory. In the special case of consciousness, however, the target phenomenon is based solely on a subject relation, that is, there can only be an intra-subjective epistemic base. As Searle puts it, "we can´t make that sort of appearance-reality distinction about consciousness because consciousness consists in the appearances themselves." (Searle, 1999, p. 456.)

This, however, does not have any "deep consequences" for the prospect of a veritable science of consciousness as far as Searle is concerned. Rather, it is a "trivial consequence of the pragmatics of our definitional practices." (Searle, 1999, p. 456,) I would suggest this is broadly consistent with what I mean when I state that the theory-based physicalism I argue for here provides a priori entailment of the target phenomenon in principle, while it does not reveal anything about experiential states in practice. It does not in any way reduce the phenomenal property associated with qualia in such a way as to diminish the validity or salience of the intra-subjective epistemic base, i.e., the knowledge in practice. I will reconvene with this point momentarily.

Moreover, as Searle admonishes, on the epistemic irreducibility of subjective consciousness, "a trivial result such as this has only trivial consequences for the unity of our overall scientific worldview." (Searle, 1999, p. 456). While stating that consciousness is irreducible in this way, the main concern for Searle is to dispel the notion of ontological

dualism in the manner that Jackson´s Knowledge Argument is seen to produce, as he also pointed out. Like Churchland argued too, as we have seen, the apparently irreducible nature of the phenomenal property of consciousness does not reveal anything about its metaphysical status. Searle would emphatically agree there is no reason to suppose some otherworldly domain to describe consciousness states.

Now, while I wish to point out how this Higher Dimensional Physicalism argued for here is consistent with the kind of emergence Searle argues, I will point out how it also goes further than the alleged irreducibility Searle describes. If Higher Dimensional Physicalism is correct, it would indeed introduce an object-relation associated with Phenomenal-Consciousness. The objective basis is the string-theoretical properties, which explain the interactions of the natural forces in general, considered as the micro-level physical properties beyond our empirical observation, from which the higher-level *physical* phenomena of consciousness states emerge. In this way, the phenomenal properties associated with qualia become reducible to string-theoretical properties *in principle* and on the fundamental level, thus considered as physical facts Mary can have knowledge about. As such, and again, if Mary knew all the physical facts, according to Higher Dimensional Physicalism, she would be able to deduce the presence of qualia in *other* minds by a priori entailment. This is the basis for the theory-based physicalism Higher Dimensional Physicalism represents.

Importantly, since I argue this reduction provides a priori entailment understood as knowledge of qualia in principle, not in practice, it does not pretend to be provide a "nothing-but" kind of reduction as Searle discusses that term. A "nothing-but" kind of reduction would be one whereby the phenomenal property is simply "nothing-but" these statements about propositional knowledge. A plump for such a "nothing-but" kind of reduction would amount to illusionism about phenomenal properties of consciousness.

By contrast, the kind of ontological reduction I argue for here, which would be broadly consistent with the ontology Searle argues for to explain consciousness as a physical phenomeon, would in no way be understood to undermine or deny the intra-subjective epistemic base of phenomenal consciousness states as they *appear* to us from the first-person perspective, which is the basis for our knowledge in *practice*. Indeed, and to reiterate Searle´s remark on the same, "No sane person could deny its existence, though many pretend to do so." (Searle, 1999, p. 457.)

Moreover, a little later in the same debate with Eccles referenced above, Searle goes on to argue against the dualist claim that the apparent causal efficacy of phenomenal properties poses a problem for materialism.

> "I do believe that thought processes have the power to operate causally, but they do that in exactly the same way that higher-level phenomena cause other, lower-level phenomena. So, for example, the heat that is generated in the cylinders of my car, that is generated by the firings of the spark plug, the heat can cause an explosion that then runs the car. Now, the heat is a higher-level property of a series of movements of electrons between the electrodes. again, that´s a banal example, let me apply that to bodily movement…" (PhilosophyOverdose, 2022, time stamp: 28:48-29:34)

He then proceeds with some remarks detailing the relevant neurochemical processes associated with motor control. Now, as I have argued, such details of NCC, if constrained by four-dimensional physicalism, do not in themselves entail the relevant phenomenal properties. These would constitute the functional, cognitive, intentional states associated with Access-Consciousness (which can of course overlap with phenomenal-consciousness). However, this what Searle says about how higher-level phenomena can cause other, lower-level phenomena is precisely how I too consider the causal efficacy of phenomenal properties of consciousness.

As I say, these string theoretical interactions are the lower-level physical features, acting upon the higher-level physical properties observed in the four-dimensional physicalist interpretation of the NCC. If we consider phenomenal-consciousness states to be the higher-level feature of the relevant string theoretical properties, there is no reason it could not produce physical effects on the lower-level physical features, similarly to how heat can ignite a fire or cause an explosion, as Searle points out.

In this way, while a dualist would have to suppose some psychophysical laws bridging two separate ontological domains to explain the apparent causality of phenomenal properties, according to higher-dimensional physicalism, the "causal efficacy" becomes tantamount to the quantum mechanical principles governing string theoretical properties in general. We should note here, however, how the notion of causality breaks down in the context of quantum mechanics. However, to the extent that these quantum mechanical principles have real world effects, which they do, the point carries force.

3.6. The Field of Consciousness

I am suggesting that consciousness is instantiated in and constituted by these higher spatial dimensions conjectured in string theory, which has been developed to entail a unified account of the electromagnetic field, the strong and weak nuclear forces, as well as gravitation. This provides a compelling and powerful view of how subjective experiential states possibly

become *realised by* specific interactions of the natural forces on a fundamental level of physical description, whereby phenomenal properties are the emergent property of these lower-level physical phenomena.

This would have deep and far-reaching consequences for how we can come to understand consciousness as fitting into Nature, perhaps in more subtle and surprising ways than I can fit in this paper. The concern here, rather, takes for its basis a view to explain the private, first-personal perspective of phenomenal consciousness. As argued in chapter 2, the phenomenal properties are not entailed by the current standard for the NCC reporting on Access-Consciousness states. Properly understood, however, on the philosophical level of analysis Higher Dimensional Physicalism can provide a theory-based framework for how the natural forces interfaces with, and emerges from, the neural architecture of a complex, self-organising biological system. It provides a compelling account of how the wet and gooey bio-physical system, as is the brain and body, can give rise to experiential states, based on natural principles, based on the interactions between the fundamental forces of nature described by physics.

On this view, the features of mind become deeply ingrained and wedded with the physical basis for descriptions of the neurobiological structures of the brain, based on natural principles. If we take seriously these features of the mind, described in terms of phenomenology on the philosophical level of analysis, these features become compatible with, and appropriately constrained by, the lower-level physical phenomena observable in a lab setting, considered as manifestations of the various interactions of the fundamental forces of nature working on the neurobiological structure of the brain. In other words, it provides a framework to provide a priori entailment of Phenomenal-Consciousness on the philosophical level of analysis, to see how the phenomenology fits with the NCC on the neurobiological level of analysis.

Understandably, this may at first seem highly abstract and idealised. However, I might go a little further in way of concretisation by considering the fundamental natural force that is electromagnetism. First, it is obvious that the mind is somehow deeply connected with the brain. As a neural substrate, the brain is an electro-chemical neurobiological structure which gives rise to consciousness states. This much is obvious, and the effects of (Access-) consciousness states are readily observable by the NCC, as discussed in chapter 2.

Electromagnetism is an integral measure for these consciousness states. Indeed, many of the different technological instruments utilised in the investigations of the NCC relies on the electromagnetic techniques for measurement of the target phenomenon. These include, for

example electroencephalography (EEG), magnetoencephalography (MEG), electrocorticography (ECoG) and transcranial magnetic stimulation (TMS) which manipulates neural activity to induce different brain states. (Mashour, Hudetz, 2018; Massimini, *et al.,* 2012; Roelfsema, *et al.,* 2018)

Indeed, in their influential paper on the prospect of a neurobiological theory of consciousness, Crick and Koch (1998) suggested that consciousness states reliant on sufficient working memory and visual system processing may be strongly correlated with semi-synchronous neural oscillations in the 40-70hz range. Whether this particular interpretation has proved lasting is beside the point. The point is the general feature of brain states from which consciousness arises being reliant on electromagnetic principles. I believe this may seem telling.

According to string theory, the electromagnetic force is itself understood as fitted in higher dimensional space. Further, the electromagnetic force can be measured and observed according to its effects in four-dimensional physics. However, consider for example the part of the spectrum of electromagnetic radiation we define as the spectrum of visible light. This obtains on the wavelengths ranging from about 750-380nm. What we perceive as red would be those wavelengths ranging from around 750-620nm. According to Higher Dimensional Physicalism, we directly perceive higher spatial dimensions from the private, first-person perspective since this is where the phenomenal properties of consciousness are instantiated and constituted. This may seem demystifying, then, of the perceptions of phenomenal redness, for example, which is simply an effect of the physical phenomenon that is electromagnetic radiation; a natural force entailed by higher dimensional space according to string theory, which we can otherwise measure empirically in the standard four dimensions of space and time.

The claim here thus becomes that the empirically observable neural substrate of the brain and the NCC are only a cross-section of mind/brain, where the mind/brain is taken as a multidimensional physical entity extending into hyperspace, to explain consciousness. The empirically observable correlates of consciousness obtain on the four dimensions of classical physics on the macro scale, whereas the phenomenal properties associated with consciousness obtain in the higher spatial dimensions, only observable from the private, first-personal perspective. These will remain empirically unobservable until we invent some supermicroscope for Mary to peer through (or some other technological tool for measurement) to observe these additional spatial dimensions and determine whether another supposed conscious system possess the relevant physical-cum-phenomenal properties

associated with qualia. Beyond propositional knowledge on the relevant string theoretical dimensions providing a priori entailment of qualia, such an instrument for observation/measurement would give her the ability to evaluate the structure and availability of phenomenal consciousness in *ex cranio* brains or silicon based conscious systems as discussed in chapter 2.

Chapter 4: Higher Dimensional Physicalism Vs. Jackson´s Epiphenomenalism

It may at first seem strange and quixotic to consider phenomenal properties as located in higher spatial dimensions beyond those we can perceive empirically as I have argued in the preceding chapter. However, to the extent that these higher spatial dimensions are taken to exist as part of physics for reasons completely independent of the question of consciousness, I would suggest it seems much less strange than dualism in general or even the epiphenomenalism argued for by Jackson more specifically.

The epiphenomenalism Jackson argues for subsequent to his Knowledge Argument will be the topic of exploration in this final chapter. In this final section of his paper, ´Epiphenomenal Qualia´(1982), Jackson argues for the causal inefficacy of the phenomenal property associated with qualia, which in themselves become understood as existing outside of physical facts of nature. In his discussion, he raises three objections to his view of epiphenomenalism and promptly seeks to refute each one in turn, in defence of his conclusion going in its favour. I will assess each of his responses to these objections in turn.

4.1. Jackson´s Epiphenomenalism

As a subspecies of dualism, epiphenomenalism supposes that phenomenal properties are distinct from physical properties in principle. To explain the relationship between them, then, we must posit some psychophysical law governing the relationship between the mind and the brain. In more general terms, Montero and Papineau provide an apt description of epiphenomenalism:

> "Epiphenomenalists view the conscious mind as an inefficacious side-effect of the brain´s operations. They agree with physicalists that bodily movements and so on are fully accounted for by brain processes, but insist that the conscious mind floats above these brain processes, as it were, rather like the puffs of smoke that are emitted by a steam train, but which themselves make no causal contribution to the train´s progress." (Montero, Papineau, 2016, p. 191.)

As such, according to the epiphenomenalism Jackson argued for, there would be a one-way causal relation from the physical to mental. The phenomenal properties associated with the features of mental life are themselves causally impotent, as Jackson puts it, but considered to be caused as a by-product of physical processes. What he is implying more specifically is a property dualism on phenomenal consciousness. What is interesting is the

way he argues towards this conclusion. He raises three objections to epiphenomenal qualia and seeks to refute them each in turn.

First is the notion that we may be wrong about the causal links determining phenomenal properties. The justification for this claim is what he takes to be the inductive basis for our inferences about causal relations. We can make note of how A is always followed by B. However, as he puts it by "reversing Hume … Anything can fail to cause anything." (Jackson, 1982, p. 133.) Despite how A always seems to be followed by B, this does not prove the validity of a hypothesis supposing a causal relation between the two. As he says, "[this hypothesis] can always be overturned by an over-arching theory which shows the two as distinct effects of a common underlying cause." (Jackson, 1982, p. 133.)

The example he uses to demonstrate this is learning that the Spurs won by reading a report about it in *The Times*. Now, it is safe to assume that this report in *The Times* means there is an equivalent report in another newspaper, *The Telegraph.* However, there is no direct causal link between the two. That is, the report in *The Times* is not what caused *The Telegraph* to write their report. There is another underlying cause which resulted in both reports, namely the fact that a football match transpired where the team from Tottenham happen to have won. We can readily understand how the report in *The Telegraph* occurred quite independently of the report in *The Times* in this way since the two had each their own representative journalists observing and reporting on the match. Though we can usually correctly infer that a report about this result in *The Times* means there will be an equivalent report in *The Telegraph,* this is not taken to imply any causal relationship between the two. Common sense knows that there is an underlying cause producing the effects that are these independent reports.

This note about the epistemic basis for causality is fine and well. Correlation is not causation. With this objection, Jackson´s aim is to refute how sensations of pain result in a bodily movement away from the source of the pain, and to extricate from the physical causal relations the phenomenal property of consciousness taken to exist as an epiphenomenon. Even still, examples like this hardly seem conclusive to support the claim, especially since he cannot point to any positive account of such an over-arching theory to show how phenomenal properties is a distinctive effect of an underlying common cause for the concomitant bodily movement, he merely hints at its possibility. Moreover, the presupposition is that the phenomenal properties exist in another ontological realm from the physical causal relations from which they are taken to arise as an epiphenomenon.

The second objection he raises is one based on Darwin´s Theory of Evolution. Evolutionary Theory states that traits are naturally selected for by environmental pressures. The traits which are conducive to survival, that optimise the fitness function, we might say, are those which will be passed on to posterity. Maladaptive traits are negatively selected for since the individual organisms who maintain these will be outcompeted by their fitter counterparts.

But then the question becomes, how come the evolutionary development of such phenomenal properties associated with these physical processes if properties are themselves causally impotent? After all, their causal impotence would mean that they do nothing to improve our fitness function in evolutionary terms, since they cannot influence behaviour in any which way. The argument he provides for considering qualia as epiphenomenal is that these result as a by-product of physical processes which developed to cause other fitness-improving functions.

In this vein, he provides an excellent analogy concerning polar bears. Polar bears developed a thick fur coat to help them withstand the cold in arctic climates. A by-product of a thick coat is that it is heavy, which would result as a non-adaptive trait that does not improve the fitness function. Indeed, in this example, the weight of the fur coat would represent a disadvantage since it debilitates movement. What Jackson is invoking here is what palaeontologist, Stephen Jay Gould and geneticist Richard Lewontin have referred to as a "spandrel" (Gould, Lewontin, 1979).

In the context of biological evolution, the notion of spandrels was introduced and developed by Gould and Lewontin to go against what they refer to us as the "adaptionist programme", which they recognised as the governing intuition among students of evolutionary theory at the time. The adaptionist programme refers to the tendency to seek explanation of all traits as evolutionary adaptations which are positively selected for to improve the fitness function of the organism. However, they argued that this at times forces an interpretation of adaptive fitness of traits which are in themselves non-adaptive. As such, a spandrel refers to a phenotypic trait which develops as the non-adaptive by-product of another adaptive trait or traits which are positively selected for to optimise the fitness function of the organism. As such, spandrels emerge as an epiphenomenon.

Thus, if phenomenal properties of consciousness are considered as epiphenomenal in this sense, from the evolutionary perspective we might consider it non-adaptive, since it does nothing to improve the fitness function. When it comes to phenomenal consciousness we would have been just as well off without it. This would be consistent with the

epiphenomenalism Jackson argues for since here the phenomenal properties associated with qualia are “causally impotent” and, thus, have no effect on the behavioural output of the organism.

However, in all other cases of such epiphenomenal spandrels I can conceive of, these would not imply any kind of dualism with the physical features, properties or causes from which they emerge as a by-product. In the case of Jackson´s epiphenomenal qualia, however, we are to believe that the phenomenal properties associated with qualia result as a by-product of physical facts about traits, features, or properties in such a way as to be somehow ontologically distinct from their causal basis, by the dualism he implies.

In the special case of phenomenal properties, then, we are to believe that the physical produces something non-physical, which remains outside of physical description even in principle. Incidentally, Montero and Papineau mentioned above also point this out,

> “Epiphenomenalism requires us to suppose that conscious states, even though they are caused by processes in the physical world, have no effects on that world. This is a very odd kind of causal structure. Nature displays no other examples of such one-way causal intercourse between realms.” (Montero, Papineau, 2016, p. 191.)

I would go so far as suggest that this reach for property dualism by Jackson reveals little more than a tacit indebtedness to the ideological grip dualism has had on our conception of the mind-brain relationship since the beginning of the modern period. It is simply a relic of dualism, to steal from Searle his point in the debate referenced in section 3.5. It admits of nothing but a “[reflection of] an antecedent metaphysical division in the furniture of the world.” As we heard Churchland call it in section 1.5.

We can see the effect of this ideological grip in Jackson´s own example with polars bears. What Jackson fails to consider with his analogy is how, in the case of the weight of the polar bear´s fur coat, or indeed with any other example of such spandrels, these are phenotypic traits which supervene on the physical traits, features, or properties from which they are caused as a by-product, and thus remain apt for physical description. The weight of the polar bear´s thick coat has mass. Further, these will often have physical effects, precisely as in the case of the heavy quality of the thick fur developed by polar bears as is Jackson´s example. The weight of this fur coat has a marked physical effect, namely that it would slow down the movements of the polar bear. Thus, the analogy with polar bears fails to show how dualistic notions of epiphenomenalism might be sustained, and Jackson´s own argument works against the conclusion he is driving towards.

Finally, the last objection he raises concerns the practical limitations of our own cognitive abilities to understand various complex phenomena. Here, I will let him introduce the basis for the claim in full himself. He states,

> "Physicalists typically emphasise that we are part of nature on their view, which is fair enough. But if we are part of nature, we are as nature has left us after however many years of evolution it is, and each step in that evolutionary progression has been a matter of chance constrained just by the need to preserve or increase survival value. The wonder is that we understand as much as we do, and there is no wonder that there should be matters which fall quite outside or comprehension. Perhaps exactly how epiphenomenal qualia fit into the scheme of things is one such." (Jackson, 1982, p. 135.)

In the first instance, I am sympathetic to this point raised by Jackson. What he correctly identifies here is the great difficulty involved in explaining how and why phenomenal consciousness arises as the result of physical processes, based on what, realistically, is the relatively modest capacity for understanding, in the grand scheme of things. This is perhaps why Chalmers (1995) appropriately formulated this question of how and why concomitant consciousness states arises from physical processes as "The Hard Problem of Consciousness."

We may also recognise here, however, how Jackson implies that the burden of evidence lies on those who would claim it is based in physical facts. This is no doubt due to the conclusion he drew based on the Knowledge Argument against physicalism, which motivates this property dualistic turn his epiphenomenalism implies.

As we have seen by the discussions in chapter 1, however, the Knowledge Argument against physicalism is wanting. Indeed, we can recall of Churchland went so far as to suggest it simply begs the question against physicalism, by exploiting the tacit and implicit equivocation of ways in which Mary knows about the facts.

Even still, it seems reasonable to suppose as Jackson does that our cognitive faculties may be too feeble to begin to understand the basis of our own consciousness. Other philosophers of mind, like McGinn has indeed argued towards this conclusion. As he had put it,

> "The limits of our minds are just not the limits of reality. It is deplorably anthropocentric to insist that reality can be constrained by what the human mind can conceive." (McGinn, 1999, p. 538.)

In this vein, it may well be that we are no more apt to discern the true nature of our conscious experience than the scientist sea slugs Jackson goes on to discuss are apt to unlock

the mysteries of the universe from their home on the seabed, as is how he concludes this section where he argues for epiphenomenal qualia.

I too often consider that we might do well to temper our hopes with a healthy dose of sobering cynicism in this regard. It is feasible that consciousness simply did not develop in our case for the purpose to envelope itself and, accordingly, does not feature the relevant faculties by which it might render itself scrutable and thus intelligible onto itself. If so, we would do well to consider that, in our search for a fundamental understanding of consciousness, and in good keeping with the name cynic, we become much like the dog chasing its own tail.

If this is the case, then no amount of our physics could provide description for conscious experience or phenomenal properties. However, this is not because it is impossible to reduce qualia to a completed account of the physical facts in principle, i.e., that physical facts are principally inadequate to entail qualia, as Jackson would have us believe. Rather, it would be due to the practical limitations of our anthropogenic imaginative and rational capacities, and the relative lack in our reasoning powers in general.

The point I wish to impress here is how such practical limitations do not imply any conclusion about how consciousness states are somehow independent of physical facts in principle, as with dualism. Consider, for instance, if we figure out a way to improve our overall level of intelligence by increasing the degree of neural plasticity, for example by boosting the availability and utilisation of brain-derived neurotrophic factor (BDNF) by the brain (or some such feat of neurobiological enhancement). Perhaps we might otherwise augment the relevant cognitive faculties to improve overall intelligence with the implementation of advanced brain-computer interface technology (Roelfsema, *et al.,* 2018). In such cases, we could conceivably increase our reasoning powers, enabling us to learn about more advanced and sophisticated physical descriptions which would entail the relevant facts about the physically constitutive basis of our experiential states (i.e., about the possible physical basis for the phenomenal properties associated with qualia). This should at least seem conceivable, also for Jackson. As such, again it seems to be case that his assertion for epiphenomenal qualia rely on little more than presupposition.

4.2. An Antidote to Jackson´s Dualism

What might we draw from Jackson´s defence of epiphenomenal qualia? As we have seen, nothing in any of his defences carry any force. In response to the first objection based on the

inductive basis for inferences of causality, he can provide no over-arching theory which might demonstrate the common underlying causal mechanism which would result in the causally impotent phenomenal property of qualia.

The same could be said in the case of his response to the third objection. Indeed, we may recognise how his suggestion in connection with the final objection, that we are ill-disposed to understand the basis for our own experiental state as such, might preclude the practical probability of us arriving at an understanding of such an over-arching theory to explain the underlying common cause between phenomenal properties and physical causal relations in general.

Further, In the case of his response to the second objection based on evolutionary theory and epiphenomenal spandrels, we can recognise how his own example works against the conclusion he drives towards. Thus, his attempt to defend a dualistic type of epiphenomenalism on phenomenal properties fails to obtain.

While I agree that it seems conceivable that we may not be equipped with the necessary cognitive capabilities to understand the basis for our own experiential states and how these fit in with the natural scheme, I do not believe this as a matter of principle. Indeed, as I have argued in the foregoing chapters, with the maturation of our understanding within modern physics as that which string theory/M-Theory represents, it can indeed be possible to fit the phenomenal properties of consciousness within the general scheme of Nature. In this way, we may finally disabuse ourselves of the dominating ideological grip dualism as had on our ways of conceptualising about the mind-body problem since at least the beginning of the modern period, when Descartes first threw down his gauntlet.

Moreover, in light of Jackson´s discussions as detailed in this chapter, I will conclude this here by suggesting to you now that the Higher Dimensional Physicalism argued for here provides a physicalist monism which would go a long way in describing such an "account of the relation between physical processes and consciousness, explaining on the basis of natural principles how and why it is that physical processes are associated with states of experience." (Chalmers, 2011, p. 3) Jackson touched upon. Incidentally, this is Chalmers´s qualification for what a meaningful response to the Hard Problem of Consciousness would need to entail.

Further, as we have seen, nothing in Jackson´s refutations of the objections against dualistic epiphenomenalism he raises provides anything with force to sustain his conclusions for its defence. By contrast, Higher Dimensional Physicalism dispels the notion of ontological dualism between mind and brain, to show how phenomenal properties can fit in a general natural scheme. In addition to providing a more scientifically based account, Higher

Dimensional Physicalism sustains the theoretical virtue of ontological parsimony whereby, according to Ockham's Razor, one ontological domain is better than two, given that it is sufficient to explain the target phenomenon. There is one point I wish to draw the reader's attention to on this score.

It has been pointed out to me that it may seem less ontologically parsimonious to introduce these additional spatial dimensions to account for the instantiation of phenomenal properties in the wider natural scheme. However, here I would point out the way in which Higher Dimensional Physicalism is a theory-based physicalism conditional on string theoretical physics. As such, these added spatial dimensions are not introduced by Higher Dimensional Physicalism itself. Rather, Higher Dimensional Physicalism relies on this theoretical framework within physics which posits the additional spatial dimensions for reasons completely independent of any theory of consciousness.

Accordingly, to the extent that string theory/M-Theory or some sufficiently similar multidimensional theory of modern physics is considered valid, these dimensions obtain in the natural scheme irrespective of any theory of consciousness. Thus, if string theory/M-Theory is correct, these added spatial dimensions would still feature in the physical ontology of Nature even given a supposed dualism between mind and body to explain consciousness. Moreover, this is the very same world in which Jackson posits his dualist epiphenomenalism to explain qualia, arguing for their causal inefficacy. By contrast, I will conclude it would be more ontologically parsimonious to fit phenomenal properties into the natural scheme based on this theory-based physicalist approach.

Concluding Remarks

With the development of my thesis argument, I have argued for what I have called Higher Dimensional Physicalism. This theory-based physicalism conditional on string theory/M-Theory, argues that we can provide a priori entailment of the phenomenal properties associated with qualia, to arrive at knowledge of the target phenomenon in principle, by understanding these as instantiated in and constituted by string theoretical properties, thus considered as physical facts. This has presented a novel refutation against Jackson´s Knowledge Argument against physicalism, which argues that phenomenal properties must be understood as existing outside of physical facts.

I have argued that the Knowledge Argument can be faulted for equivocation on the relevant ways Mary knows about the facts between the two stated premises, to help invalidate its conclusion. However, to the extent that our investigations into the neural correlates of consciousness rely on empirical observations of neurobiological mechanisms obtaining in four dimensions of space and time, these can only entail reports on Access-Consciousness states. Thus, any theory on the neurobiological level of analysis constrained by what I have referred to as Four-Dimensional Physicalism will remain inadequate to provide a priori entailment of the target phenomenon, to include Phenomenal-Consciousness, and the Knowledge argument against physicalism is sustained.

By contrast, according to Higher Dimensional Physicalism, the as of yet empirically unobservable string theoretical properties become the micro-level physical properties from which higher-level physical-cum-phenomenal properties such as Phenomenal-Consciousness states emerge. This reveals a viable physicalist monism consistent with scientific ontology, to explain how phenomenological states on the philosophical level of analysis fit in with investigations on the NCC on the neurobiological level of analysis.

Further, the private nature of phenomenal properties as it seems to us from the first-personal perspective can be explained by their instantiation in higher spatial dimensions which are themselves beyond empirical observation, where they become understood as realised by the fundamental forces of nature, most notably electromagnetism, working on the neural substrate of the brain. If further developed, this can go a long way in describing how and why physical processes are associated with experiential states, on the basis of natural principles.

Bibliography and References

Ajina, S., & Bridge, H. (2017). Blindsight and Unconscious Vision: What They Teach Us about the Human Visual System. *The Neuroscientist*, *23*(5), 529–541. https://doi.org/10.1177/1073858416673817

Aru, J., Suzuki, M., & Larkum, M. E. (2021). Cellular Mechanisms of Conscious Processing: (Trends in Cognitive Sciences, 24, 814-825; 2020). *Trends in Cognitive Sciences*, *25*(12), 1096–1096. https://doi.org/10.1016/j.tics.2021.09.008

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron, 76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038.

Bayne, T., Hohwy, J., & Owen, A. M. (2016). Are There Levels of Consciousness? *Trends in Cognitive Sciences, 20*(6), 405-413. https://doi.org/10.1016/j.tics.2016.03.009

Bayne, T., Seth, A. K., & Massimini, M. (2020). Are There Islands of Awareness? *Trends in Neurosciences (Regular Ed.)*, *43*(1), 6–16. https://doi.org/10.1016/j.tins.2019.11.003

BBC News. (2023, January 17). Nick Cave Says ChatGPT's AI Attempt to Write Nick Cave Lyrics 'Sucks.' *Entertainment & Arts*. https://www.bbc.com/news/entertainment-arts-64302944.

Bender, E., Mcmillan-Major, A., Shmitchell, S., Gebru, T., & Shmitchell, S.-G. (2021, March). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3442188.3445922.

Bickle, J. (2020). Multiple Realizability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition). URL: https://plato.stanford.edu/archives/sum2020/entries/multiple-realizability/

Block, N. (1995). On a confusion about a function of consciousness. *The Behavioral and Brain Sciences*, *18*(2), 227–247. https://doi.org/10.1017/S0140525X00038188

Blumenhagen, R., Lüst, D., & Theisen, S. (2012). String Dualities and M-Theory. In *Basic Concepts of String Theory* (Theoretical and Mathematical Physics). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-29497-6_18

Bourget, D., & Mendelovici, A. (2019). Phenomenal Intentionality. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition). URL: https://plato.stanford.edu/archives/fall2019/entries/phenomenal-intentionality/

Chalmers, D. J. (2010). Consciousness and Its Place in Nature. In *Oxford University Press eBooks*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195311105.003.0005

Chalmers. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.
Chalmers, D. J. (2020). What is conceptual engineering and what should it be? *Inquiry (Oslo)*, *ahead-of-print*(ahead-of-print), 1–18. https://doi.org/10.1080/0020174X.2020.1817141

Churchland, P. M. (1999). Knowing Qualia: A Reply to Jackson. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 571-578). The MIT Press.

Churchland, P. M. (1985). Reduction, Qualia, and the Direct Introspection of Brain States. *The Journal of Philosophy*, *82*(1), 8–28. https://doi.org/10.2307/2026509

Churchland, P. S. (1986). *Neurophilosophy: toward a unified science of the mind-brain* (pp. XI, 546). MIT Press.

Crick, F., & Koch, C. (1998). Towards a Neurobiological Theory of Consciousness. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 277–291). The MIT Press.

Dehaene, S., Changeux, J.-P., & Naccache, L. (2011). The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications. *Research and Perspectives in Neurosciences*, 55–84. https://doi.org/10.1007/978-3-642-18015-6_4.

Del Pin, S. H., Skóra, Z., Sandberg, K., Overgaard, M., & Wierzchoń, M. (2021). Comparing theories of consciousness: why it matters and how to do it. *Neuroscience of Consciousness*, *2021*(2), niab019–niab019. https://doi.org/10.1093/nc/niab019

Doerig, A., Scharnowski, F., & Herzog, M. H. (2019). Building perception block by block: a response to Fekete et al. *Neuroscience of Consciousness*, *2019*(1), niy012. https://doi.org/10.1093/nc/niy012

Doerig, A., Schurger, A., Hess, K., & Herzog, M. H. (2019). The unfolding argument: Why IIT and other causal structure theories cannot explain consciousness. *Consciousness and Cognition, 72*, 49-59. https://doi.org/10.1016/j.concog.2019.04.002

Dupré, J. (1993). *The disorder of things: metaphysical foundations of the disunity of science* (pp. VIII, 308). Harvard University Press.

Flanagan, O. (1999). Prospects for a Unified Theory of Consciousness or, What Dreams Are Made Of. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 97-110). The MIT Press.

Forrest, P. (2020). The Identity of Indiscernibles. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition). URL: https://plato.stanford.edu/archives/win2020/entries/identity-indiscernible/

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

Friston, K. (2005). A Theory of Cortical Responses. *Philosophical Transactions: Biological Sciences, 360*(1456), 815–836. http://www.jstor.org/stable/30041303.

Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *205*(1161), 581–598. https://doi.org/10.1098/rspb.1979.0086

Graziano, M. S. A., Guterstam, A., Bio, B. J., & Wilterson, A. I. (2020). Toward a standard model of consciousness: Reconciling the attention schema, global workspace, higher-order thought, and illusionist theories. *Cognitive Neuropsychology*, *37*(3-4), 155–172. https://doi.org/10.1080/02643294.2019.1670630

Greene, B. (n.d.). String Theory - Dimensions and Vibrations. *Encyclopedia Britannica*. https://www.britannica.com/science/string-theory/Dimensions-and-vibrations.

Gregersen, E. (2018, October 19). How Albert Einstein Developed the Theory of General Relativity. *Encyclopedia Britannica*. https://www.britannica.com/story/how-albert-einstein-developed-the-theory-of-general-relativity.

Gulick, R. V. (1999). Understanding the Phenomenal Mind: Are We All Just Armadillos? Part: Phenomenal Knowledge and Explanatory Gaps. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 435-442). The MIT Press.

Güzeldere, G. (1999). The Many Faces of Consciousness: A Field Guide. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 1-67). The MIT Press.

Hawking, S., & Mlodinow, L. (2011). *The Grand Design: New Answers to the Ultimate Questions of Life.* Bantam Books.

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, *1*(II). https://doi.org/10.33735/phimisci.2020.II.64

Hurler, K. (2023, March 15). Chat-GPT Pretended to Be Blind and Tricked a Human into Solving a CAPTCHA. *Gizmodo*. https://gizmodo.com/gpt4-open-ai-chatbot-task-rabbit-chatgpt-1850227471.

Jackson, F. (1982). Epiphenomenal Qualia. *The Philosophical Quarterly*, *32*(127), 127–136. https://doi.org/10.2307/2960077

Jackson, F. (1999). Finding the Mind in the Natural World. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 1-67). The MIT Press.

Jackson, F. (1999). What Mary Didn´t Know. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 483-492). The MIT Press.

Jefferson, G. (1949). The Mind Of Mechanical Man. *The British Medical Journal, 1*(4616), 1105–1110. http://www.jstor.org/stable/25372573.

Jolij, J. (2022). A Higher Dimension of Consciousness: Constructing an empirically falsifiable panpsychist model of consciousness. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2204.07164

Kaku, M. (1994). *Hyperspace: a scientific odyssey through parallel universes, time warps, and the tenth dimension* (pp. XVI, 359). Oxford University Press.

Kaku, M. (2005). *Parallel worlds: a journey through creation, higher dimensions, and the future of the cosmos* (pp. xvii, 428). Doubleday.

Khalil, S., Kane, G., Antoniadis, I., & Moretti, S. (2016). Supersymmetry, Supergravity, and Superstring Phenomenology. *Advances in High Energy Physics*, *2016*, 1–1. https://doi.org/10.1155/2016/3595120

Kiefer, C., & Weber, C. (2005). On the interaction of mesoscopic quantum systems with gravity. *Annalen Der Physik*, *517*(4), 253–278. https://doi.org/10.1002/andp.20055170404

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews. Neuroscience*, *17*(5), 307–321. https://doi.org/10.1038/nrn.2016.22

Kripke, S. A. (1999). The Identity Thesis. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 445-450). The MIT Press.

Lamme, V. A. (2004). Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. *Neural Networks*, *17*(5), 861–872. https://doi.org/10.1016/j.neunet.2004.02.005

LeDoux, J. E., Michel, M., & Lau, H. (2020). A little history goes a long way toward understanding why we study consciousness the way we do today. *Proceedings of the National Academy of Sciences - PNAS*, *117*(13), 6976–6984. https://doi.org/10.1073/pnas.1921623117

Lewis, D. (1999). What Experience Teaches. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 579-596). The MIT Press.

Lycan, W. (2023). Representational Theories of Consciousness. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition). Forthcoming. URL: https://plato.stanford.edu/archives/win2023/entries/consciousness-representational/

Lycan, W. (2023). Representational Theories of Consciousness. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition). Forthcoming. URL: https://plato.stanford.edu/archives/win2023/entries/consciousness-representational/

MacDonald, A. A., Naci, L., MacDonald, P. A., & Owen, A. M. (2015). Anesthesia and neuroimaging: investigating the neural correlates of unconsciousness. *Trends in Cognitive Sciences*, *19*(2), 100–107. https://doi.org/10.1016/j.tics.2014.12.005

Mandler, G. (1975). Consciousness: Respectable, useful, and probably necessary. In R. L. Solso (Ed.), *Information Processing and Cognition*. Lawrence Erlbaum.

Mashour, G. A., & Hudetz, A. G. (2018). Neural Correlates of Unconsciousness in Large-Scale Brain Networks. *Trends in Neurosciences (Regular Ed.)*, *41*(3), 150–160. https://doi.org/10.1016/j.tins.2018.01.003

Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron (Cambridge, Mass.)*, *105*(5), 776–798. https://doi.org/10.1016/j.neuron.2020.01.026

Massimini, M., Ferrarelli, F., Sarasso, S., & Tononi, G. (2012). Cortical mechanisms of loss of consciousness: insight from TMS/EEG studies. *Archives Italiennes de Biologie*, *150*(2/3), 44–55. https://doi.org/10.4449/aib.v150i2.1361

McGinn, C. (1999). Can We Solve the Mind-Body Problem?. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 529-542). The MIT Press.

Mills, E. (1996). Interactionism and Overdetermination. *American Philosophical Quarterly*, *33*(1), 105–117. http://www.jstor.org/stable/20009850

Montero, B. G., & Papineau, D. (2016). Naturalism and physicalism. In *The Blackwell Companion to Naturalism* (pp. 182-195).

Mørch, H. (2024). *Non-physicalist Theories of Consciousness* (Elements in Philosophy of Mind). Cambridge: Cambridge University Press.
(Reviewed copy of pre-print by access from author)

Nagel, T. (1998). What Is It Like to Be a Bat? In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 519–527). The MIT Press.

Narain, K. S. (1989). New Heterotic String Theories in Uncompactified Dimensions < 10. *Current Physics–Sources and Comments*, January, 246–251. https://doi.org/10.1016/b978-0-444-87492-4.50023-9.

Northoff, G., & Huang, Z. (2017). How do the brain's time and space mediate consciousness and its different dimensions? Temporo-spatial theory of consciousness (TTC). *Neuroscience and Biobehavioral Reviews*, *80*, 630–645. https://doi.org/10.1016/j.neubiorev.2017.07.013

OpenAI. (2023). *Review of GPT-4 Technical Report.* ArXiv. https://cdn.openai.com/papers/gpt-4.pdf.

Oppy, G., & Dowe, D. (2021). The Turing Test. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition). URL: https://plato.stanford.edu/archives/win2021/entries/turing-test/

Overgaard, M. (2017). The status and future of consciousness research. *Frontiers in Psychology*, *8*, 1719–1719. https://doi.org/10.3389/fpsyg.2017.01719

Papineau, D. (2002). *Thinking about consciousness* (pp. XIV, 266). Clarendon.
[PhilosophyOverdose]. (2022, December 17). Debate on Mind-Brain Relation: Searle vs Eccles (1984) [Video]. YouTube. https://www.youtube.com/watch?v=1KemhfmAsg8

Poincaré, H. (1913). The Relativity of Space. *The Monist, 23*(2), 161–180. http://www.jstor.org/stable/27900426.

Putnam, H. (2011). Minds and machines. In *Philosophical Papers, Volume 2* (Vol. 2). Cambridge University Press.

Roelfsema, P. R., Denys, D., & Klink, P. C. (2018). Mind Reading and Writing: The Future of Neurotechnology. *Trends in Cognitive Sciences*, *22*(7), 598–610. https://doi.org/10.1016/j.tics.2018.04.001

Searle, J. R. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Searle, J. R. (1999). Reductionism and the Irreducibility of Consciousness. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 451-460). The MIT Press.

Sjöstedt-Hughes, P. (2022, April 14). Consciousness and higher spatial dimensions. *IAI TV*. https://iai.tv/articles/consciousness-and-higher-spatial-dimensions-auid-2107?_auid=2020

Skirry, J. (n.d.). René Descartes: The Mind-Body Distinction. *Internet Encyclopedia of Philosophy*. ISSN 2161-0002. Retrieved from http://iep.utm.edu/rene-descartes-mind-body-distinction-dualism/

Smolin, L. (2022). On the Place of Qualia in a Relational Universe. In *Consciousness and Quantum Mechanics*. Oxford University Press. https://doi.org/10.1093/oso/9780197501665.003.0018

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*(11), 1–29. https://doi.org/10.1037/h0093759

Stoljar, D. (2023). Physicalism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition). URL: https://plato.stanford.edu/archives/sum2023/entries/physicalism/

Strawson, G. (2006). Realistic Monism: Why Physicalism Entails Panpsychism. *Journal of Consciousness Studies*, *13*(10), 3–31.

Strawson, P. F. (1974). Self, Mind and Body. In *Freedom and Resentment: And Other Essays* (pp. 169-178). Methuen.

Swaab, D. (2014). *We Are Our Brains: From the Womb to Alzheimer's*. Penguin UK.

Tononi, G., & Koch, C. (2015). Consciousness: here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1668), 117–134. https://doi.org/10.1098/rstb.2014.0167

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *59*(236), 433–460. http://www.jstor.org/stable/2251299

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need*. arXiv preprint. https://arxiv.org/abs/1706.03762

Zimmerman, D. (2010). From Property Dualism to Substance Dualism. *Aristotelian Society Supplementary Volume, 84*, 119-150. https://doi.org/10.1111/j.1467-8349.2010.00189.x