COPYRIGHT AND CITATION CONSIDERATIONS FOR THIS THESIS/ DISSERTATION

- **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

- **NonCommercial** — You may not use the material for commercial purposes.

- **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

**How to cite this thesis**

Bifactualism

A Physicalist Account of Experience

By

Danielle Swanepoel
Abstract

Philosophy of mind has begun to rely on input from neurobiology and neuroscience to answer questions concerning consciousness, representation and the subjective character of experience. Some philosophers believe that through studies done on the brain, neuroscience will help us answer the hard-problems of consciousness. The first chapter of this paper is concerned with the kind of contributions neurobiology can make to certain debates in philosophy of mind and proceeds to explain that even though neurobiology is mostly a positive contributor to philosophy of mind, it still fails to answer some of the more pressing issues in philosophy of mind.

In the second chapter of this paper, I focus on Frank Jackson’s Knowledge Argument. Jackson’s Knowledge Argument is perhaps one of the most challenged arguments concerning experience in philosophy of mind and responses to this argument will possibly tell us more about the mind. Jackson argues that we gain a new kind of knowledge when we experience something, for example when seeing the colour red. He argues that these new learnt facts are non-physical. David Lewis argues that a person learns no new facts upon experiencing red, but rather abilities such as the ability to recognize, imagine and remember. In this chapter I also examine some of the counterarguments to Jackson’s Knowledge Argument.
and conclude that these philosophers have approached the Knowledge Argument incorrectly. I suggest a different physicalist response to the Knowledge Argument.

In the third chapter of this paper, I propose a new physicalist account of experience I call ‘bifactualism’. The Knowledge Argument is an argument for dualism that claims that there are both physical and non-physical facts which can be learnt through experience. I reject the Knowledge Argument and suggest bifactualism. Bifactualism is a new physicalist account essentially comprising two elements. First, it distinguishes between two kinds of physical facts: general and particular. The second element is the claim that this distinction explains facts about consciousness, without resorting to dualism. I disagree with the dualist claims made in the Knowledge Argument and show that the Knowledge Argument neither supports dualist claims nor does it refute bifactualist claims.

Most contributions made with regards to the Knowledge Argument focus on what Mary is able to learn once outside the black and white room. Bifactualism is interested in what she is able to learn in the black and white room which makes this a novel approach to the Knowledge Argument.

In the fourth chapter of this paper I propose bifactualist responses to several issues that have been highlighted throughout this paper. In this chapter, I primarily focus on Nagel’s What it is Like to be a Bat? Nagel claims that we cannot know what it is like to be a bat because we can never have access to its
subjective character of experience. This chapter argues that it is as difficult to know the feeling of what it is like, or WIL (Prinz, 2012), to be a bat which has a feeling of WIL, as it is to know the feeling of what it is like to be a book (which has no feeling of what it is like: non-WIL). I argue that this is not because of two different ways of knowing two different properties, but rather that there are two different physical facts about both WIL and non-WIL properties. I show that with a bifactualist account, there are particular physical facts that can be known about WIL and non-WIL properties alike that are not expressible in the language of physics.
Contents

Introduction ................................................................................................................................... 8

Chapter 1: Neurobiological Contributions to Philosophy of Mind ........................................... 14

1. Philosophical Insinuations .................................................................................................. 15

2. NCC: An Unsatisfactory Explanation ............................................................................. 17

3. Cognition as an Embodied Process .............................................................................. 21

4. Consciousness: Generated by the Body ...................................................................... 26

5. The Different Levels of Consciousness ...................................................................... 40

6. Why the Different Levels of Consciousness Hypothesis is Problematic .................. 46

7. Is Felt-value and Somatic-data Enough for Consciousness? ..................................... 52

8. Mental Maintenance ....................................................................................................... 58

9. Where Neurobiology Fails ........................................................................................... 61

10. Conclusion ...................................................................................................................... 66

Chapter 2: What Experience Tells us about the Mind .......................................................... 69

1. What Experience Tells Us ............................................................................................... 70

2. Frank Jackson: The Knowledge Argument ................................................................... 71
3. David Lewis: The Ability Hypothesis .................................................... 73
4. Michael Tye: A Challenge to Both Jackson and Lewis ......................... 77
5. Coleman: Why the Ability Hypothesis is Problematic .......................... 80
6. Brian Loar: Phenomenal States .............................................................. 82
7. Benjamin Kuipers: Complex Information Processing .......................... 84
8. Churchland: Knowing-by-Acquaintance .............................................. 85
9. Why the Knowledge Argument is Problematic ..................................... 89
10. Conclusion .......................................................................................... 94

Chapter 3: Bifactualism: A Physicalist Account ......................................... 98

1. The Importance of Facts .................................................................... 99
   A) Bifactualism: A Physicalist Account ............................................ 100
2. General and Particular Facts are Physical ............................................ 106
   B) Applying Bifactualism to Consciousness ...................................... 120
3. Physical Facts in Jackson’s Knowledge Argument .............................. 120
   3.1 Bifactualism and Learning of Facts ............................................ 127
4. Conclusion .......................................................................................... 129

Chapter 4: The Bifactualist Response to the Mind-Body Problem ................ 132
1. Three Contributions to the Mind-body Problem ................................. 133
2. Neurobiology vs. Bifactualism.............................................................. 134
3. The Knowledge Argument vs. Bifactualism ....................................... 139
4. Nagel on the Problem of Access .......................................................... 141
5. Access to WIL and Non-WIL ............................................................... 144
6. The Bifactualist Response to the Problem of WIL and non-WIL ....... 146
7. What MaryActually Learns ................................................................. 149
8. Red: According to John ...................................................................... 151
9. Conclusion .......................................................................................... 154

Conclusion .............................................................................................................. 157

Reference List ........................................................................................................... 160
Philosophy of mind concerns itself with concepts such as (amongst others) consciousness, identity, mental properties and their relation to the physical or non-physical and the existence and composition of mind. One of the central themes in philosophy of mind is the debate between dualism and physicalism. Dualists claim that mental states or mental phenomena are non-physical and are not identical with the body. Physicalists claim that all that can be said of the mind can also equally be said of brain activity or the body. Physicalists argue that everything that exists is physical. Reductive physicalism argues that mental states can be reduced analytically to a physical basis, and thus that all facts about mental states can be expressed in the language of physics. Non-reductive physicalism claims that mental states are constituted by physical states; however these mental states are not analytically reducible to the physical. Functionalism, a branch of physicalism, would make the claim that the mind is what the brain does and therefore is nothing other than neuronal activity in the brain.

Several arguments in philosophy of mind attempt to prove either dualism or physicalism correct. Even though it seems some variety of physicalism is the most
plausible option in current philosophy of mind\textsuperscript{1}, there are still serious questions or issues that remain unresolved and these issues at times favour physicalism and other times, dualism. Amongst others, issues in philosophy of mind include questions about phenomenal states, qualia, consciousness, awareness, sense of self, and identity. The existence of the above issues seems to make dualism an attractive option at times. This is because the difficulty with relying on empirical data to support these issues in philosophy of mind is that currently, there is not a lot of empirical evidence concerning the nature of these kinds of states or properties.

Finding consciousness in neuronal activity is an ongoing challenge and understanding the anatomy of phenomenal states or qualia presents unanswered questions. How one is aware or is able to own an identity of self is a concept the physicalist could attempt to answer by appealing to studies done on the brain, because neuroscience and neurobiology have shown that certain characteristics of consciousness and self can be affected through various brain lesions or other such damage. However, whatever the findings, dualists could claim that they still do

\textsuperscript{1}Physicalism seems more plausible at present because our scientific enquiry is such that it supports physical findings. This has got to do with the manner in which scientists enquire about the mind by studying or analysing the structure and function of objects which are measurable or analysable from an empirical analysis. If scientific studies are to be done on the mind, then the results will naturally be physical.
not quite answer the harder-problems of consciousness and of mind which I will discuss in chapter one: \textit{Neurobiological Contributions to Philosophy of Mind}.

Using neuroscience and neurobiology to answer questions in philosophy of mind may not appeal to many proponents of dualism because it would seem that the only answers to come from these endeavours would be materialist or physicalist in nature. This is so because of the physical nature of this scientific enquiry. The problem dualists face arises from the causation issue: how does something non-physical interact and have causal relations with something physical? Studies in neurobiology and neuroscience deal only with the physical aspect of things and therefore the issue of interaction or causation that the dualist proposes will not really feature in scientific studies. Even if it did feature in their studies, it still remains doubtful that physical enquiry will completely satisfy dualist claims (see footnote 1). However, I would like to caution proponents of physicalism to also be wary of relying on neurobiology or neuroscience to answer questions in philosophy of mind. Some findings in neurobiology and neuroscience point to physicalism being right, but sometimes, \textit{not finding} something inadvertently supports some dualist claims. An example of this would be the lack of physical evidence of phenomenal states or qualia\textsuperscript{2}.

\footnote{\textsuperscript{2}“The term ‘qualia,’ as standardly used, refers to the properties of mental states that characterize what it is like to have them” (Chalmers, 2002: 198).}
Some studies in neurobiology and neuroscience have shown that there is indeed value in the human being possessing phenomenal states and qualia and this means that physicalists may need to re-examine some of their strongest claims, one claim from certain physicalist camps being that qualia do not exist. Also, if neurobiology can find the feeling of what it is like in the brain, it will lead to certain dualist camps having to redefine the idea of a mind distinct from the body given how central that is to some dualist arguments. Neurobiology may also show that the ‘mind’ is dependent on or even formed in both the body and the brain. This would mean that proponents who claim that the mind can only be found in the brain (which allows for the plausibility of the brain in a vat example), may need to re-evaluate their claims or refine their arguments. Remember that for some dualists, the mind and body are separate substances or things which are able to interact and have some causal effects on each other. Some physicalists, in particular Central State Materialism, will argue that the brain is the mind etc. If we make the claim that the mind is found within both body and brain, then dualists cannot propose separate substances and this will support the physicalist account as the physicalist may not be as concerned with knowing where exactly the mind is, as long as it is somehow established that it is physical.

3 Later, I discuss the idea that the body and brain belong to the same system, but I cannot do away with referring to them as two separate structures that work together. When I use the term body and brain, I need the reader to keep in mind that this distinction is not necessary. Both are, according to certain neuroscientists, integral to the forming and functioning of the mind.
In the first chapter of this paper, I look at how some theorists in neurobiology and neuroscience address the concept of mind as an embodied process. In particular I look at Antonio Damasio’s work to support the hypothesis of embodied cognition. I argue that even though neurobiology can show us where neural correlates of consciousness can be mapped as well as how it is possibly generated, it still does very little in explaining what it feels like for us to be conscious or how we should go about explaining consciousness in terms acceptable to both a physicalist and dualist. I show how neurobiology fails to answer certain philosophical questions.

Once I have shown how neurobiology cannot give a satisfactory account of the certain philosophical concerns, I move onto the second chapter of the paper where I discuss Frank Jackson’s Knowledge Argument. Jackson argues that there are non-physical facts to be learnt about the world and he is able to make this claim based on a contentious thought-experiment about Mary, a neuroscientist, who is locked in a black and white room her entire life. According to Jackson, Mary learns all physical facts in the black and white room and when she is released from the room, she learns non-physical facts. I discuss this argument in-depth as well as several counterarguments or contributions made to this debate. I show how the Knowledge Argument is also unable to answer the more pressing issues in philosophy of mind.
It seems, then, that both attempts to buttress physicalism by appeal to neurobiological explanations of the mind, and dualists such as the Knowledge Argument, face persistent challenges. Against this background, I propose a physicalist position called bifactualism: a bifactual account of experience. Bifactualism does two things: firstly it redefines the kind of physical facts we can have access to, in a way that concedes something to the Knowledge Argument without giving way to dualism. Secondly, it offers a possible account of consciousness.

Through bifactualism, I show that the Knowledge Argument is lacking in two ways: clarity with regards to physicality of information and a proper exposition as to what Mary is able to know in the black and white room. Bifactualism contributes to both these aspects and suggests that Mary’s knowledge is incomplete in the black and white room, thereby challenging Jackson’s first premise which most theorists choose not to confront. After defining and explaining the bifactualist view and its contribution to the Knowledge Argument, I offer a tentative bifactualist response to the mind-body problem.
Chapter 1
Neurobiological Contributions to Philosophy of Mind

Some philosophers and certain fields in philosophy are cautious with regards to contributions made by science, and perhaps especially with regards to neuroscience. It is debateable as to whether or not neuroscience or neurobiology can add much value to debates in philosophy of mind. In this chapter I look at contributions from neuroscience and neurobiology regarding consciousness and mind. This chapter deals with certain concepts such as Neural correlates of consciousness (NCC), embodied cognition, and consciousness as it is understood by several neuroscientists (Damasio, Morrison, Tversky etc). For some, the mind does not exist only in the brain but is brought about through an embodied process and subsequently gives rise to an embodied cognition. In this chapter I discuss both NCC and embodied cognition and conclude that neurobiology certainly has a lot to offer but still fails to address certain pressing issues in philosophy of mind.
1. Philosophical Insinuations

Several scientists and philosophers have looked to the brain and its workings to answer questions with regards to the mind and consciousness. What exactly the role of the brain is regarding the ability for it to engender feeling, emotion, thoughts, memories and even animal/instinctive consciousness is still a mystery. Some researchers believe that finding the neural correlates of consciousness (NCC)\(^4\) will answer many of the as yet unanswered questions in philosophy of mind concerning consciousness; these questions concern (amongst others) the placing of the mind, identity, free-will, acquisition of knowledge etc. But perhaps we are not in the position to assume that the brain is able to do all or indeed any of these things. It is important to remember that there are philosophical arguments purporting to show that consciousness may not be the neuronal activity itself but rather that certain neuronal activity brings about or is accompanied by some form of consciousness or conscious experience. We will consider some of these arguments in later chapters.

Recently, neurobiology and neuroscience have embarked on studying consciousness as a process potentially involving the whole body rather than limiting empirical consciousness-study to study of the brain. This embodied

\(^4\) The Neural Correlates of Consciousness (NCC) is defined as a minimum set of the neuronal activities sufficient for a conscious experience (Tononi and Koch, 2008: 239).
process is such that the brain and the body are both responsible for the making of consciousness and thus the making of the mind. This embodied process could be considered as the body feeling and being aware of itself as such. The consequence of understanding the mind in this way is that NCC, even if properly mapped out, will not tell us much about what this embodiment feels like. In fact, neuroscience and neurobiology are quick to use the term ‘feeling’ in discussions about embodied cognition without realising that there are philosophical insinuations to this kind of claim. Physicalists find it difficult to pin-point exactly what this feeling is within a physicalist account and dualists claim that this feeling is non-physical.

Neurobiological contributions will be the basis of this chapter and I refer to Damasio throughout my paper as my reference for most of the discussions on neurobiology because he is a neuroscientist who has written extensively on the subject of consciousness. His work represents a reasonably well-developed example of attempts by neuroscientists to shed light on the nature of consciousness through both hard empirical research and more philosophical speculation inspired by that research. One of the central themes Damasio concentrates on in his various works is the idea of feeling consciousness which is important this chapter.
The chapter will firstly discuss why the concept of NCC fails to capture the complexity of consciousness. Secondly, this chapter will discuss the idea of embodied cognition and through discussing several ideas Damasio puts across in his work, we will understand how this embodied cognition is established or generated. Thirdly, we look at the philosophical understanding of the mind and consciousness to show how this comprehensive study of an embodied cognition may bring us no closer to understanding consciousness. We will conclude that neurobiology may have successfully shown us how consciousness can be generated, but it does very little in helping us understand exactly what consciousness is or how we are to begin accessing and studying it.

2. NCC: An Unsatisfactory Explanation

Some philosophers (e.g. Crick and Kock, 2003) believe that finding the NCC will offer several answers to some of the most pressing philosophical issues. Other philosophers (e.g. Noël and Thompson, 2004) do not believe that finding the NCC will tell us much more than where the activity of consciousness can be found in the brain. Moreover, it may not show where consciousness can be found but rather that certain kinds of neuronal activity give rise to consciousness or the appearance of consciousness. This does little more, I believe, then show us that somehow consciousness is linked to neuronal activity.
Finding the NCC may not show us the true nature of subjective experience, introspection, sense of self or even how the qualitative character of experience is engendered or represented in the brain. The reason that finding the NCC may not show any of this lies in our inability to access these states. These states are first-person qualitative experiences and presumably, prima facie, inaccessible from third-person enquiry. This inaccessibility of subjective experience from a third-person perspective will be discussed in more detail in the chapter: *The Bifactualist Response to the Mind-body Problem.*

Several neuroscientists and philosophers believe that finding the NCC will somehow disprove at least substance dualism and strengthen the physicalist position. However, it is not clear that finding the NCC would disprove property dualism. Substance dualism would need to show how a causal relationship can exist between the NCC, which are physical, and the mental substance, which is difficult if they are distinctly different substances. Property dualism can admit that NCC may show a causal relationship between mental and physical states, but this does not prove that mental properties are physical properties.

The substance dualist makes the claim that no one substance can have both mental and physical properties (Rodriguez-Pereyra, 2008: 69), and therefore that mental properties need to be of a different substance than physical properties. Even though these substances are distinctly different, they must be able to interact with
one another, assuming that there are causal relationships between mental and physical states. This is problematic. How is it possible for a physical substance to interact and have a causal relationship with a non-physical substance, given that they can have none of the same properties?

Burge (2007: 344) argues that some philosophers “are concerned whether the intentional ‘aspects’ or qualitative ‘aspects’ of mental events are epiphenomenal—that is, lacking in causal power and irrelevant to causal transactions”. To assume that mental events are somehow causes is to assume that they are physical. If mental events are epiphenomenal, then we cannot discuss any causal interactive role without answering the interaction problem. Burge (2007: 344) claims that physical states are usually considered as generating or “underlying” mental states and suggests that this kind of token-identity theory\(^5\) of mental states supports materialism and in this case, the physicalist account.

Physicalists dismiss this predicament by suggesting that causal relations can only occur between two physical properties and if there is any causal relation occurring between the mental and physical, then they must both be physical properties. If

\(^5\)Token-identity theory asserts that “every type (i.e. kind) of mental state is identical with some type of bodily (primarily neurophysiological) state”. An instance of this would be to maintain that every pain is identical with C-fiber firing or simulation. Davidson’s version of Token-identity theory is anomalous monism which rejects this kind of relation. “Instead it is proposed that every token (i.e. particular instance) of a mental state is identical with some token or bodily state” (Mautner, 2005: 294).
physicalists are correct, one would assume that all properties being physical, neurobiology would somehow have made a bigger impact on studies in philosophy of mind and satisfactorily answered the causation and interaction problem.

One of the ways that neurobiology can contribute to this debate is by showing that mental properties are physical properties and therefore examinable through neurobiological enquiry. This has not happened yet. Some philosophers believe that this will never happen. Currently, another less contentious contribution neurobiology can make to the mind problem is to show a correlation between states and events in the mind and the brain-body. However, finding NCC is a way of viewing the results of these relations; it will not show us the correlated event itself. Neural correlates are such that some neuronal activity has a correlate ‘episode’ of consciousness. NCC does not decisively tell us about the relationship between mental events and the brain. A correlation shows a kind of causal relationship but cannot decisively show us the exact properties of this ‘episode’. It can be understood that NCC can show a constitutive or dependency relation of the mental on the physical and Burge (2007: 358) claims that “It is not obvious, however, that this dependency need involve material constitution—the mental events being identical with or made up of physical events”.


According to Damasio, finding the NCC will in actual fact not prove much in the way of knowing how consciousness works or how it is formed or brought about. For Damasio, the entire body makes up consciousness and mindfulness and it is not exclusively housed within the brain (Damasio, 1999). Therefore, NCC do not answer the mind problem for Damasio. If consciousness is experienced in the body and through the body, then mapping NCC will not explain how consciousness works.

I argue that knowing where to find something does not mean that one knows how it works. A simple example is the car that you drive. For the general driver, we know where the engine is and we know it needs fuel etc, but this does not mean we know exactly how the engine is able to propel the car forward or backward (unless you are a mechanical engineer and for the sake of this example I hope you are not). Even if consciousness has neural correlates, it cannot explain how consciousness is created. Therefore, finding NCC will fail to answer one pertinent question concerning consciousness: 1) how exactly is this consciousness generated?

3. Cognition as an Embodied Process

In attempt to answer the question of how consciousness is generated, I will devote the next section to a discussion on embodied cognition. Even though finding NCC is a neurobiological endeavour, scientists working within the theoretical
framework of embodied cognition are not as concerned with finding neural correlates but are rather interested in how consciousness and mind are formed as an embodied process. As one of these scientists, Damasio does not necessarily seem as interested in finding NCC either, but rather how the brain and the body work as an integrated system in forming consciousness. For Damasio, finding the NCC will not answer the “hard problems” of mind. However, neurobiology may contribute by showing causal dependencies between events in the mind and the brain. If neurobiology can show causal relations between the mind and the brain, then substance dualism would need to explain how it is possible that this relationship takes place.

The concept of embodied cognition and Damasio’s theory of consciousness holds some significance for philosophy of mind. Embodied cognition is defined by Anderson (2005: 2) as “the body [which] has a special status in and for cognition at several organizational and structural levels. It is not only neurons (or sub-neuronal structures) that matter; nor is it only the interaction of organism and environment. Rather, structure and function, action and interaction, matter from top to bottom, affecting the nature and content of mental entities and events”.

Proponents for embodied cognition believe that consciousness is a process which

---

6 Hard Problems are what Chalmers would consider the problem of experience (2002: 198). How we feel when we experience something cannot be explained by science as easily as some of the easy problems which include motors skills and basic awareness.
is brought about by the workings of the body as an entirety and Damasio goes about trying to show how this consciousness is generated through processes of the body. Even though the concept of embodied cognition is framed within a neuroscientific/neurobiological framework, there are several philosophical insinuations (intentional or unintentional) in the theory, and in particular Damasio’s work, which I hope to explore.

Tversky and Hard (2009) and Damasio (1994) are interested in how consciousness is formed in the human being. Tversky and Hard are primarily interested in cognition as an embodied process and Damasio is interested in consciousness as an embodied process. In this paper, I do not wish to argue the difference between cognition and consciousness but rather assume that they both encompass the sense and awareness of self. Some may argue that they are distinct because cognition seems to be aimed outward in the knowing of objects outside of ourselves while consciousness seems to be directed inward in the act of knowing ourselves. I do not want to discuss this apparent difference as I think it irrelevant to the purpose of this paper. Cognition of your own body and consciousness of your own body seems to result in the same sort of awareness; highlighting any possible distinguishing features may result in a loss of focus for this paper. I would rather like to consider cognition and consciousness as similar
and generated from the embodied processes. I consider embodied cognition as being cognitive and conscious of oneself and of our external environment.

Theorists in traditional cognitive science suggest that cognitive processing occurs at a central point rather than generated or embodied in features and aspects of the physical body (Bechtel, 2009). Recently, the idea of a central viewer has been losing favour in both cognitive science and neuroscience; rather an embodied cognition is what is now being advocated. Several theorists hold that the body occupies several roles in cognition which include the body as a constraint, the body as a distributor and the body as a regulator (Borghi and Cimatti, 2010). The body as a constraint indicates the idea that cognition is limited to the features and functions of the body. Tversky and Hard (2009: 124) claim that “the mind is locked in a body that, at any time, occupies a specific place and faces a specific direction”. The first idea I would like to establish is that mindfulness (and what I take to be the embodied cognition) is very much an embodied process.

Lakoff and Johnson (1999) write that a “mind is not merely embodied, but is embodied in such a way that our conceptual systems draw largely upon the commonalities of our bodies and the environments we live in” (Lakoff and Johnson, 1999: 6). A mind “locked” in a body is an interesting concept which I would like to explore in more detail here. For the purpose of this paper, I would like to interpret this “locked-in” state as resulting from the body as a constraint. I
am not as interested in the body as distributor or the body as regulator for the
purpose of this paper. The idea that the body acts as a constraint on cognition is
supported by two different views, one being that embodied cognition is dependent
on the importance of action and the role that different bodily-states play in
cognition; and the other, that the embodied cognition is understood as a situated
cognition (again, consider this in light of the body as a constraint).

According to a radical version of embodied cognition, “cognition is constrained
by the specific kind of body we possess” (Borghi and Cimatti, 2010: 763).
Cognition is constrained in that we understand our own bodies, and objects
outside of ourselves, from a specific point of view. In order to interpret sensory
input and effectively respond with the necessary or appropriate output, we must
undergo a process of bodily situatedness in a given experience.

This situatedness is necessary for us to develop a sense of a body we possess.
Borghi and Cimatta (2010) consider this body-sense as “the very basic feeling that
I am the body that does and is in control of what this same body is actually doing
and perceiving” (Borghi and Cimatta, 2010: 767). The doing and perceiving body
is the body of which we are aware and this awareness is not a third-person,
objective awareness, but rather a specific, particular kind of awareness. Borghi
and Cimatta (2010: 768) discuss how exactly this sense of body is brought about.
It appears that the self or bodily-perception is developed from lower forms of the
body to higher forms. The most basic form is what Morin (2006) labels the “introceptive system” which are the feelings that “represent a sense of the physiological condition of the entire body” (Craig, 2003: 500). Borghi and Cimatta mention bodily perception resulting from lower forms to higher forms and Damasio suggests the same process in his own hypothesis of consciousness. I will look specifically at Damasio to explain how this consciousness supposedly surfaces.

4. **Consciousness: Generated by the Body**

At times, this discussion may seem steeped in neurobiological jargon, but the neurobiological background is important in order to better understand how Damasio reaches his conclusions, what exactly embodied cognition is supposed to be, and how it is brought about. Damasio introduces several levels or layers of consciousness and this has provoked criticism with regards to the “problem of integrating the information processed by different regions of the brain [which] is known as the binding problem” (Cleeremans 2003: 1).

Peter Lipton (1998: 212) distinguishes two kinds of binding problems: perceptual and conceptual. The perceptual binding problem is “the problem of explaining in neurophysiological terms how it is that we see the various perceptual qualities of a physical object, such as its shape, colour, location and motion, as features of a single object” and how the brain manages to bind all of these aspects of
perception together to form one unitary presentation of the object. The conceptual binding problem relates to the study of consciousness within an inter-disciplinary environment and binding ideas from across disciplines such as between neuroscience and philosophy. In this paper I am interested in both, but primarily in a different kind of binding problem between the different levels of consciousness that Damasio proposes. If it is difficult to explain how an image with “various perceptual qualities” is bound and perceived as a unitary object, then how does Damasio hope to correlate different levels of consciousness?

Damasio is not even suggesting that consciousness is generated in various regions of the brain only, but rather that this consciousness is generated from several regions of the body and encompasses different layers or levels of consciousness. How exactly are we to understand a binding between different levels of consciousness generated from different levels of the body?

Damasio proposes three levels of consciousness and the first level is created or formed from the body-proper and understood, or at least processed in the brain. Damasio understands an organism as the body-proper and the nervous system. Brown and Stenner (2009: 131) write that the “brain is part of the body-proper of a higher animal [and that] Damasio gives the brain its body back”. This would mean that the brain belongs to the body-proper but not so of the nervous system. Damasio discusses the idea of the body-proper as being the central object on
which all maps in the brain are formed by the nervous system. All subsequent levels of consciousness are based on this primary level.

Two issues arise here with regards to Damasio’s idea of consciousness: Firstly, how exactly this lower-level consciousness is formed is not certain and this is problematic as the question that arises is how can lower level events be conscious? Secondly, does Damasio help himself to the notion of lower-level consciousness to validate his claims of higher level consciousness? In an attempt to answer the first question, Damasio may claim that each level provides some kind of information to the next. I call this information somatic-data and it is this somatic-data that is the content of consciousness, not the physical processes themselves. It would be difficult to imagine a PH balance being conscious or aware. However, it seems more plausible that the somatic-data that PH balance provides can be understood within consciousness.

In an attempt to answer the second question, I am going to suggest that we give Damasio his claim to the lower-level consciousness so that he can explain the higher levels of consciousness. I suggest this for the simple reason that any kind of explanation of how consciousness comes about is better than no discussion at all and if empirical enquiry can show results regarding higher-level consciousness, then proving lower level consciousness does not seem as urgent. It
may be the case that by assuming lower-level consciousness, studies of higher-
level consciousness will yield more results.

Damasio suggests that sensory input and signals from internal regulatory
functions within and from the body are monitored in the brain, thus the brain
receives its lower level somatic-data for consciousness-development from the
body itself. Somatic-data is not to be confused with sense-data. Sense-data could
be considered as “an immediate object of sensory perception” (Mautner, 2005:
567). An example of sense data is to see a red tomato and the image forming in
one’s mind as round and red. The mental image resulting from this experience of
the tomato is considered the sense-datum (Huemer, 2011). This is different from
somatic-data.

Somatic-data is sensory and experiential information gleaned and received from
the body-proper and is not dependent (or at least not entirely so) on objects
perceived or experienced. In addition, an organism is directly aware of sense-data
and if there were no awareness, this sense-data would not exist as such. Somatic-
data on the other hand can be generated by the body-proper but only the more
important information collected will become available to awareness to the
organism. An example would be if an organism was uncomfortably hot,
uncontrollably thirsty and a little hungry. There may be a good chance that the
organism will not become aware of the hunger until the other conditions have
been satisfied. However, this does not mean that the body was not generating the somatic-data related to hunger most of the time. It is roughly as if that somatic-data becomes overtly knowable on a need-to-know basis. The question of how this somatic-data is generated or interpreted can better be understood by looking to the brain and its primary function.

The primary function of the brain is to regulate the body in its interaction and experience with the environment. This is important to understand for further discussions. The brain is a processor of sorts and processes what the body is and does through its experiential instances. Experience is an important aspect in this process. We will be looking at the philosophical issues concerning experience in the chapter: *What Experience Tells Us About the Mind*. However, let it suffice to state that for Damasio, processes which can be felt are experienced in and by the body.

Before discussing this further, let us review the context in which Damasio makes his claims. Damasio is assuming that events studied by and comprehensible to neurobiology can and do somehow give rise to consciousness. His findings are empirical in nature, but his conclusions often go somewhat beyond his data, and their statement may involve a degree of conceptual analysis. Whether or not neurobiological processes do give rise to consciousness cannot definitively be proven with the empirical data he provides. Rather I suggest we should see his
more speculative remarks as seeking to provide a conceptual framework that shows how it is possible for the events, processes and properties studied by neurobiology somehow to give rise to consciousness. Understanding his contribution as including some conceptual analysis is important.

Let us now set out the neurobiological ideas explored by Damasio and other neuroscientists. This is necessary in order for us to better understand how Damasio has reached the conclusion that consciousness is generated from the body and that there are several levels of consciousness. According to Damasio (1999), the body and brain should be considered as one combined structure which is made up of a body proper and a nervous system of body and the brain. The body and brain are connected together by neural and biochemical circuits. The purpose of these circuits is to carry signals from every part of the body to the brain and from the brain to every part of the body. According to Damasio, this circuitry is imperative to the functioning and content of the mind.

For Damasio, the circuitry of the nervous system ensures a homeostatic process within the body which regulates and monitors internal systems. Not only does this relationship ensure interaction with the external environment but it also manufactures internal responses which in turn make up somatic-data in the form of visual, audio and sensory somatic-data. The feeling of the body and how these feelings are formed in the brain are what Damasio calls, images. I feel that the
word ‘image’ is confusing in that it is immediately associated with visual perception. I have chosen to refer to these ‘images’ as *somatic-data* and will continue doing so throughout this paper. This somatic-data which is provided by the body is correlated with an assigned value and this value is *felt*. Because the word ‘feeling’ can be confused with emotion, I call the feeling of this somatic-data: *Felt-value*. Importantly here, the brain is primarily responsible for mapping (recording) this somatic-data, but this somatic-data is based on the body, without the body to provide this content, this brain remains ‘blind’ to this internal environment of the body.

This internal environment, which is supposed to be a *felt* physical process, is called (amongst other things) the *Internal Milieu*. The concept of an internal milieu was first coined by Claude Bernard (Sabbatini, 1998, 14) and Damasio uses it often throughout his work. The internal milieu is the collaboration of all the biochemical processes occurring within the organism as a whole. An analogy might be drawn with the car that you drive. Not only does the engine propel the car forward and backward because of its various parts (pistons, rings, cogs, belts etc) that work so well together, but also that all these parts and their workings are somehow monitored by electrical wiring which sends ‘signals’ to your dashboard display telling you when the oil is low or when your car is due for its next service. The internal milieu is finely tuned to the processes of the body and any deviation
from this precise range (the perfect homeostatic range) may result in illness or death. The precise range mentioned is functioning that occurs at a cellular level which is finely tuned according to its temperature, PH levels, chemical balances etc. all for the smooth and controlled running of the organism.

The internal milieu functions at a physical-chemical level and according to Damasio the awareness of this internal milieu also takes place in the form of emotions and feelings. Damasio’s (1994, 139) point is that it is imperative to think of emotions and feelings not as something belonging to a non-physical realm but rather as a physical occurrence that takes place through an internal milieu as well as neural categorizations within the brain. Other than it being a physical occurrence, these emotions and feelings are also vital in our receiving this data and interpreting it. Within the brain itself, somatic-data is gathered, formed and recorded with regards to the organism and its environment.

Remember that this somatic-data encompasses all cellular processes, bodily-experiences or sensory occurrences the human body has. Therefore, experience is very important for the individual to learn about herself (body, brain and sense of self and her environment).

Damasio’s different levels contributing to consciousness are supposed to explain the functioning of mind from the lowest bodily level (instinctive or animal bodily functions, such as PH balance and Basal Temperature) to that of the feeling of
what it is like (a higher awareness of self and of one’s surroundings). Remember that the PH balance and other animal bodily functions also form a part of the foundation that contributes to the somatic-data bank. Damasio claims that the most basic ingredients for minds being conscious are wakefulness and images (somatic-data) (Damasio, 2012: 186). According to Damasio, the human brain has the ability to create “maps”. These maps are the way the brain “informs itself” (Damasio, 2012: 63). Mosca (2000: 8) states that “all maps are neural maps and all can become mental images”. Along with creating maps, the brain also generates somatic-data (images) which, for Damasio, is the main foundation of mind.

The particular role of consciousness is to experience the neural maps as somatic-data (remember that this is data which is felt), to be able to interpret the somatic-data as well as to act accordingly. Just to reiterate, somatic-data is the information gained from the body’s internal and external environment. Somatic-data must be formed in order to develop and establish the mind and must also be experienced in order to form consciousness. According to Damasio (2012: 63), maps are formed when the body interacts with its environment and maps can also be constructed when the brain recalls somatic-data from its memory. For Damasio, it is almost as if the brain represents or imitates “aspects of the structure of nonbrain things and events, which includes the actions carried out by our organism and its
components, like limbs, part of the phonatory apparatus, and so forth” (Damasio, 2012: 64). This somatic-data represents all aspects of the organism experiencing itself and its environment.

This would entail some kind of resemblance relation between the brain and the rest of the body. Somatic-data is a form of representation similar to nomic covariation. Representationalists believe that “mental representations are variously semantically evaluable” (Maloney, 1994: 241) and nomic covariation accounts for the notion of “carrying information” (Gomila, 1994: 605). An example Gomila uses is if $x$ carries information about $y$, then according to its nomic relation, $x$ and $y$ somehow covary. Fodor (1981) and Dretske (1983) would agree that this entails that $x$ means $y$. However, the issue would be trying to find out what exactly $x$ is to mean $y$. Internalists would argue that “the phenomenal character of experience is entirely a matter of the representational content of the experience” (Thompson, 2008: 398) and not of the external properties of the object that is being experienced as externalists would argue. Internalists hold that the representation quality is not reliant or dependent on objects or the external world but rather on the what it is like content of a person seeing a red tomato. There is no external content being represented as an object but rather the way my experience represents the object. Somatic-data would be more of an internalist representation.
Damasio suggests that the neurons in the brain, when firing\(^7\), form actual patterns which represent aspects of the body as well as the external environment. Again, these patterns of neuronal activity should not be thought of as a visual representation. Rather, it is an informational representation of the body. Damasio explains that after cutting parallel to the brain surface and extracting a thin slice of brain, one would see a “sheathlike structure that essentially resembles a two-dimensional grid, the main elements in the grid are neurons, displayed horizontally” (Damasio, 2012: 66). He suggests that the lines that make up the grid are formed by the neurons themselves and that when neurons are firing, a pattern is formed quite distinct from the neurons that are not firing. This does not mean that the “distinct pattern” actually “mirror” or directly imitates the external environment. However, it is possible that Damasio can here be misinterpreted.

\(^7\) Neurons are the cells necessary for satisfactory brain function. The neurons (of which there are billions) are connected as intricate circuitry which is aligned in layers. Collections of neurons may not necessary have a strict format or structure and can take the form of a bundle or random group. There are three main parts that make up a neuron. A cell, an output fibre (axon) and several input fibres (dendrites). When neuroscientists refer to a neuron firing, what they are actually referring to is the way a neuron communicates with the neurons around it. In its active state, a neuron is known to ‘fire’. When this firing occurs, the cell body sends an electric current down the axon which is then transferred to the synapse. The synapse allows for certain chemicals (also known as neurotransmitters) to be released and these neurotransmitters operate as receptors (Damasio, 1994: 28-29).
Damasio can be misunderstood when he suggests that neurons actually generate some form of pattern. R.W Sperry (1952) in his article: *Neurology and the Mind-Brain Problem* discussed what may seem to be a similar idea of a triangle being depicted within the brain when a triangular shape was viewed. He called this doctrine *psychoneural isomorphism* and comments that “isomorphism...represents the most natural and naive way of dealing with mind-brain relations” (Sperry, 1952: 284). Daniel Siegel (2007: 185) also discusses a neural map as being a cluster of firing neurons that represent the object it is mapping. What exactly they mean by this form of representation is difficult to define. Psychoneural isomorphism is not what Damasio is suggesting. Damasio is improving upon the notion of psychoneural isomorphism by arguing that the neurons do change shape or structure in order to imitate their environment but not in the way that isomorphism suggests.

A simple example of how unsound isomorphism is would be to consider neurons turning blue when we see the colour blue. A more absurd example (if possible) would be a neuron sounding out a sharp sound when we hear a police siren. Sperry (1952) considers this as a naive interpretation and so would Damasio. Bennett and Hacker have misinterpreted Damasio by thinking he is committed to psychoneural isomorphism. Bennett and Hacker’s (2003: 139) rebuttal to Damasio’s idea of map-forming is that not all that one perceives with one’s
sensory organs is an object, or that not all a person perceives can be translated into somatic-data. With regards to the latter concern, this somatic-data is information about the body interacting with the environment. While I agree with Bennett and Hacker that not all we perceive are objects, I do not know why somatic-data cannot represent many states of the body, regardless of whether or not an object is being perceived or one is imagining an object which is not actively being perceived at that moment.

Also, the mention of “sensory organs” is problematic. I do not think Damasio has committed himself to suggesting that all somatic-data is dependent on what is being perceived by sensory organs. I interpret Damasio as suggesting that any process in the body can in principle be mapped as somatic-data, not only those processes occurring in the sense organs. The somatic-data is information of the body and is felt. This is a very crude explanation, but if each neuron is felt, then neurons that cluster together are all felt in a certain way as a whole episode or event. Also, regarding the former criticism from Bennett and Hacker’s who argue that not all that is perceived is an object or can be translated into somatic-data, I would expect that Damasio would argue that these processes of perception (concerning all the senses) are not about objects or things but about our own bodies in relation to these things.
Here it is important to remember that the mapping process is not limited to visual stimuli but rather mapping takes place with regards to all sensory input and output. This will include the somatic-data, and consequent mapping, of the body itself viewing or perceiving. The somatic-data that are formed is not a visual representation of the body but are rather a felt representation of the body sensing an external as well as an internal environment. For Damasio, this somatic-data and map-forming is where or how consciousness is generated\(^8\). Similar to the idea of somatic-data, Oliver Sacks (1986) claims that there is a “secret sense or sixth sense” which is “the continuous and unconscious sensory flow from the movable parts of our body (muscles, tendons, joints), by which their position and tone and motion is continually monitored and adjusted, but in a way which is hidden from us because it is automatic and unconscious” (1986:47). This “sixth sense”, or proprioception, that Sacks mentions is unconscious in the sense that we are unaware of it at the lower levels.

Sacks (1986: 47) explains that proprioception is an “unconscious sensory flow from the movable parts of our body (muscles, tendons, joints) by which their position and tone and motion is continually monitored and adjusted, but in a way that is hidden from us because it is automatic and unconscious”. Sacks discusses

---

\(^8\) Again, Damasio’s work must be interpreted within a conceptual framework because he suggests this hypothesis but does not have sound empirical evidence to substantiate this claim.
Sherrington’s coinage of this term in the 1890s and explains that proprieception is different from ‘extroception and introception’ because it is necessary for “our sense of selves; for it is only by courtesy of proprioception, so to speak, that we feel our bodies as proper to us, as our ‘property’, as our own” (Sacks, 1986: 47). Importantly, proprieception results from the movable parts of the body and somatic-data supersedes this as it also able to provide information of all bodily states. Somatic-data is generated from lower-level bodily states and is made available as overt awareness on a need-to-know-basis. Somatic-data are an interesting contribution to the notion of embodied cognition and help us better understand how this embodied cognition is generated.

5. The Different Levels of Consciousness

Damasio mentions four levels of consciousness\(^9\) (1999), but the first three levels of consciousness seem to be the most prevalent and most widely discussed amongst critics and supporters of his work. The three levels are the protoself, the core-self and the autobiographical self. Damasio extensively writes about different levels of consciousness and how they are possibly generated, but in my studies of Damasio I found that a major issue remains unanswered: I do not think

\(^9\) The fourth level of consciousness is called the meta-self and is the level of consciousness which is able to “view” the other levels of consciousness. The issue arising from this will be discussed in the paper.
Damasio adequately discusses how awareness is engendered in neuronal activity. He claims that consciousness is understood through neuronal activity but does not quite explain how one becomes aware of this consciousness he proposes. In this chapter I am giving as faithful an account of Damasio’s work as I can as an illustration of the sort of thing neuroscientists explore in their consciousness studies. This does not however mean that I think Damasio has all the answers. However, what he proposes is interesting and conducive for the purpose of this paper.

The protoself, according to Damasio (2012: 190) is an incorporated collection of neural maps within the brain which maps the most fundamental and basic aspects of the physical body. Damasio suggests that what makes the protoself maps unique are that not only is there somatic-data present, but also feelings of the somatic-data. These feelings, which Damasio (2012: 190) calls “primordial feelings” are present in the awake and aware brain. These protoself maps are maps depicting both what occur in the body at a very basic level and sensory portals aimed at the external environment. These maps represent basic things such as threats to the physical body which will result in a feeling of hunger, thirst or pain etc.

We briefly discussed the notion of representation in the previous section of this chapter and the question of how exactly these maps represent any states, events or
objects are not thoroughly discussed in Damasio’s work. However, I am helping myself to the notion of representation (which arguably exists more widely in the natural world) in order to get further with the notion of consciousness (which arguably is not as widespread as representation). In the previous section I suggested that somatic-data be regarded as an internalist interpretation of representation, however Damasio may not agree. It is difficult to consider Damasio’s work in light of representation given that his work is primarily concerned with consciousness and how consciousness is generated. Damasio, I believe, helps himself to the notion of representation when he suggests his different levels of consciousness.

The second level of consciousness (second-order maps), which arises from the protoself is the core-consciousness. This level of consciousness, according to Damasio (2012: 200), is where the process of constructing a self actually begins. In order to construct a self, a protagonist needs to be introduced to the equation. This protagonist is first and foremost formed by the protoself’s felt, physical body and primordial feelings mentioned above. Damasio suggests that the primordial feelings are “likely the foundation of the material me, and are, in all probability, an important and peak manifestation of consciousness in numerous living species” (Damasio, 2012: 202).
Damasio introduces a protagonist and this can be confusing. He seems to be suggesting that neurobiologically we are systems of bodily-processes or functions. Each level of consciousness contributes to the next, but there is not one overall viewer or ‘I’. The use of the word protagonist can cause this kind of confusion. Sometimes Damasio’s choice of experimental language in explaining neurobiological events or properties can be confusing. Let us, for the purpose of this paper, assume that the idea of a protagonist refers to an owner of the body and mind. This owner is a processor of functions. By this I mean that an owner is able to know, interpret and react – very much process (overtly and covertly) - functions and processes of the body.

The “second-order maps” (Siegel, 2007: 185) arise from the protoself and body’s moment-to-moment engagement with itself and the external environment. For Damasio (2012: 202), the core-self comes about when a change occurs with the protoself which instigates a sequence of events. According to Damasio (2012: 203) the first part of the sequence is a “transformation in the primordial feelings that results in ‘a feeling of knowing the object’, the feeling that differentiates the object from other objects of the moment”. Again, think of this as a processor of systems or bodily structures which is able to identify the feelings of the somatic-data.
The second part to the sequence is the consequence of knowing the said feeling. I prefer to think of these feelings as the person experiencing (feeling) the somatic-data as exclusively ours so as to cause less confusion. The idea of a feeling being owned is brought about by feelings of ownership, of something that belongs to me. Damasio (2012: 203) refers to this consequence as attention; this is the attention which is now focused on the object itself. For Damasio, the core-self is formed by linking the modified protoself to the object which is now connected to a feeling of knowing and “enhanced by attention”. So to put it simply, when a person is aware of an object, she is aware through having a feeling which is associated to that somatic-data of perceiving that object. Exactly how this is accomplished is not clear and Damasio admits a lack of scientific data in this regard. Accordingly, every object should evoke some form of feeling unique to that object or to objects of similar value. Her awareness of an orange (the fruit) will have a different feeling than a feeling she would experience when seeing an orange car. Once she has the feeling, she is then able to recognise or become (at times) aware of this feeling. Again, this feeling is allocated to somatic-data which hold a certain felt-value.

It seems that the entire process of forming both the protoself and the core-self is accompanied by the forming of somatic-data, which, according to Damasio, is paramount to producing the conscious mind. The somatic-data is also formed as a
series of events, the first being the formation of somatic-data as an organism (protoself), the second is the formation of an “object-related emotional response (that is, a feeling); and images [somatic-data] of the momentarily enhanced causative object” (Damasio, 2012: 203). From the development of the protoself and the core-self, an organism will feel a sense of ownership of the somatic-data as it is formed. This sense of ownership of the somatic-data, in the form of felt-value, is central to the sense of self.

The autobiographical self is the culminating result of the contributions of the protoself and the core-self. Simply put, the autobiographical self are “autobiographies made conscious” (Damasio, 2012: 210). Siegel (2007: 185) suggests that these “third-order maps” map the changes and movement of the core-consciousness over time. The autobiographical self is the self over the past, present and future which includes memories, ambitions, predictions, hopes and dreams etc. The autobiographical self has the unique ability to rely on the data provided by the protoself and the core-self to use the information of the past in relation to the future and to analyse how it relates to our current experiences or the present moment.

According to Damasio (2012: 212), the autobiographical-self is formed by firstly grouping together biographical memories in order to be considered as individual objects. Each individual object has the ability to modify the protoself and in-turn
engenders feelings of knowing and attention in the core self. Secondly, these said objects need to be coordinated. This coordination process must be able to call up and deliver memories etc, and the autobiographical self relies on the delivery of numerous core-self contributions that can be comprehended as momentary or transitory (Damasio, 2012: 212). The autobiographical self understands core-self contributions in the here and now, and with this is able to coordinate this information into past and future episodes.

Damasio understands these levels of consciousness as developing from the protoself to the autobiographical self. Exactly how Damasio thinks awareness is engendered in neuronal activity is still not clear. Conceptually, Damasio is suggesting that a person’s protoself collects somatic-data relating to basic bodily functions. The core-self is built upon this foundational self. The core-self provides the ownership of somatic-data collected. Finally, the autobiographical self is able to view, interpret and manipulate the somatic-data by creating a sense of self with a past, present and future. Exactly how awareness is brought about, or created, is not clear.

6. Why the Different Levels of Consciousness Hypothesis is Problematic

With the three-level consciousness Damasio proposes, it may at first glance seem as if he is victim to the Cartesian theatre mentality, three-fold. It could be said of Damasio’s notion of consciousness that the protoself is a ‘viewer’ of bodily
functions and draws maps of what it views, secondly, the core-self ‘views’ the protoself and finally the autobiographical-self is witness to its two subordinate levels of consciousness. This theory may fall victim to the notion of the homunculus which assigns every person a viewer over all systems and processes of the body. Bennett and Hacker (2003: 234) write that:

the conception of mental images as private, inner pictures that only the subject can see is the perceived conception amongst neuroscientists. Mental images [somatic-data] are commonly thought to be copies of antecedent impressions. These images [somatic-data] are supposed to be such as can be scanned, examined and turned around, and their features, like the feature of a picture, may be discerned or overlooked. Again, this conception recapitulates the errors of Descartes.

Here, I would like to discuss the error Descartes supposedly made in his own work and then relate it to what Damasio is trying to achieve with his three-consciousness proposal. The error Descartes (1637) seemingly committed was to argue that there was a viewer of sorts, a substance quite different from the body, or at least something which is somewhat different and which is able to interact with the substance of which the body is made. Also, this ‘I’ separate from the body is able to interact with the body and view the body as a kind of theatre:
I thence concluded that I was a substance whose whole essence or nature consists only in thinking, and which that it may exist, has need of no place, nor is dependent on any material thing, so that ‘I’ that is to say, the mind by which I am what I am, is wholly distinct from the body, and is even more easily known than the latter, and is such, that although the latter were not, it would still continue to be all that it is. (Descartes; 1968: 44)

For Damasio, Descartes’ idea is false because for Damasio the ‘thinking’ is clearly reliant on “the structures and operations of being” (Damasio, 1994: 248). According to Damasio, a Cartesian theatre is not what is occurring between the different levels of consciousness. There is not a separate entity or substance (a homunculus) which views the processes of the body as something separate from ‘itself’. Damasio (2012: 214) mentions the possibility of misunderstanding the concept of coordinators and identifying coordinators as possible members of the Cartesian theatre. He assures that this is not the case and suggests that there are no consciousness centres or interpreters of the different levels of consciousness, but rather that these coordinators are “spontaneous organizers of a process”. (Damasio, 2012: 214).

From Damasio’s work, it may seem that awareness is amassed in different stages of processing. What exactly is aware is discussed with the understanding that the self is a kind of an interpreter, if you will. This still does not quite get away from what the Cartesian theatre is suggesting. Even if Damasio claims that there is no
‘I’ that is consciousness of a self, I do not think he satisfactorily shows how this is not the case. Conceptually he may argue that there is no ‘I’ and that we are “operations of being” but this still does not explain how this awareness is possible. In addition, it would seem that these “operations of being” still give rise to a sense of a privileged ‘I’ and exactly how this is so, is also not adequately explained.

I think that we still view ourselves in this world as an ‘I’ and we cannot just disregard this kind of viewer of self. Even though Damasio has not satisfactorily argued for an “operations of being”; this does not mean that his ideas are completely irrelevant to philosophy of mind. Daniel Dennett (1995: 3) asks of “who or what is the audience for this "display" of "images" [somatic-data]? Not a Cartesian ego or self, isolated in some central module--the dread Cartesian Theatre--and overburdened with powers and responsibilities, but a self distributed throughout the body, a clear descendant of the Aristotelian vegetative soul”.

Dennett sees Damasio’s consciousness as a self which is not a viewer, but rather a distributed “awareness” of the human process.

How exactly this is accomplished given that we still seem to be aware or have a sense of a knower is puzzling. Although, being aware or having a sense of a knower in ourselves does not mean there is a viewer but rather the processes are experienced in a way that has one form of awareness. Empirically, Damasio
cannot prove that this process occurs. A very simple example of this “knower or viewer” would be that of a soccer team. Each player of the soccer team is a complete bundle of processes on their own and when all players come together they make a team and they play as a team. They are each aware of themselves as making up a whole and this wholeness is realised on the field. This would mean that each player would appear to make up consciousness as a whole. There is a sense in which the team is conscious in that there is a kind of team spirit or team dynamics going on which allows each individual to contribute to something which is larger than themselves. Thus, it can be said that the team is conscious of a larger goal and work together towards this shared goal, very similar to the consciousness Damasio proposes.

This analogy may seem similar to Ned Block’s discussion of Homunculi-headed Robots (1978) where he proposes a thought-experiment of the Chinese system. This experiment involves convincing the Chinese government to “realise a human mind for an hour” (1978: 96) by supplying each person of the one billion who reside there with two-way radios. Each person is responsible for performing one task in the form of either an input or output ‘signal’. He calls this the Chinese system. This system is able to process information and responds accordingly, thus forming what may seem to be a kind of consciousness. However, Block argues that realising a conscious episode does not amount to consciousness. All persons
in the Chinese system appear to satisfy something that looks like consciousness but this does not mean that any form of consciousness is taking place. Block is an anti-functionalist who argues that even if something plays a functional role, it does not mean that it is conscious. The soccer-team example may display consciousness but this does not mean that the consciousness actually exists.

I argue that this consciousness is not only possible but also quite real for one simple reason: every Chinese individual has their own consciousness which, when pooled together with all other individuals, does result in a kind of localised consciousness. The players in the soccer team also have individual consciousnesses which contribute to an overall kind of consciousness. However, an overall knowing of processes may not be possible in this regard. As pointed out before, Damasio also does not quite answer the ‘knower’ question sufficiently.

Damasio briefly explores a fourth level of consciousness called the meta-self. Dennett (1995) discusses the possibility that Damasio need not incorporate another level of consciousness such as the meta-self to complete the process of consciousness. He suggests that Damasio may have gone “off the rails just a bit at one point” (Dennett, 1995: 4) and that perhaps Damasio should reconsider his hypothesis of a meta-self. I agree with Dennett that the hypothesis of a meta-self is unnecessary with regards to knowing that one knows. If Damasio’s thesis is
correct, then the autobiographical self should suffice with regards to this because the autobiographical self has full access to the preceding forms of consciousness and is enough as an overarching processor. If the meta-self had proven to be necessary then Damasio might have actually slipped deeper into the trap of the Cartesian theatre because then a *knower of knowing* would have to have been incorporated. Fortunately, Damasio does not explore the concept of the meta-self in his latest work (2012), therefore doing away with the knower of a knower dilemma.

7. Is Felt-value and Somatic-data Enough for Consciousness?

For Damasio and other embodied cognition theorists, the importance of the allocation of felt-value to somatic-data seems to be a necessary process for consciousness-generation. Here, I briefly discuss the idea of consciousness as a physical process and *the feeling of what it is like* as being foundational for a sense of self. Understanding consciousness as a physical process is difficult. John Searle (2007) discusses the hypothesis of “basic facts” which examines the idea of all things in the world being made up of particles that “are known with certainty, in the sense that the evidence is now so great that it is irrational to doubt them” (Searle, 2007: 27). However, the question Searle is led to ask is “how can we square this self-conception of ourselves as mindful, meaning-creating, free, rational etc., agents with a universe that consists entirely of mindless,
meaningless, unfree, nonrational, brute physical particles?” (Searle, 2007: 5). Several philosophers (e.g. Descartes, Jackson) would be averse to assigning concepts such as free-will, subjectivity and meaning-creation to purely “brute physical particles”.

Damasio would argue that we allocate value to certain “brute physical particles” and it is through this felt-value that we are able to know objects and things in the external as well as internal environment. The value assigned to biological organisms or inanimate objects which are important to the person will differ greatly from value assigned to biological organisms or inanimate objects that are not important to a person. A high or strong felt-value could be assigned to a biological organism or an inanimate object, for example money, only for what that inanimate object will offer the subject.

Damasio (2012:46) further explains the concept of biological value which holds that all organisms will assign value to every item in its environment (briefly discussed before). The level of value assigned to some items and the lack of value assigned to others will demonstrate the importance of that specific item to the organism. Value, for the organism, can be assigned in several different ways. According to Damasio (2012: 47), neuroscience has recognised certain chemical molecules (dopamine, neropinepherine, serotonin, cortisol, oxytocin, and vasopressin) that are connected to the idea or concept of reward and punishment,
which for Damasio is directly associated with the notion of value. “Value relates directly or indirectly to survival. In the case of humans in particular, value also relates to the quality of that survival in the form of well-being” (Damasio, 2012: 48). Assigning felt-value relies on what can be considered advantageous for the individual, which is a very real and subjective measure; these criteria could range from something as fundamental as food to deciding to purchase an expensive sports car. Assigning felt-value relies on what Damasio calls the somatic-marker hypothesis (Damasio, 1994: 173).

The Somatic-marker hypothesis is a system of somatic-data (we have already discussed the importance of somatic-data-formation through a mapping-process in the brain) that are accompanied by feelings. Somatic-data, “especially valuable images [somatic-data], given their importance for survival, were ‘highlighted’ by emotional factors. The brain probably achieves this highlighting by generating a felt state that accompanies the image in a parallel track. The degree of feeling serves as a ‘marker’ for the relative importance of an image [somatic-data]” (Damasio, 2012: 175).

For Damasio, this “marker” does not have to be experienced as overt but can even be something as subtle as a “covert, emotion-related signal” of which the organism might not even be aware. This notion of markers is applicable to all stages contributing to consciousness, from the first-order protoself to the third-
order autobiographical self. Damasio (2012: 76) suggests that the importance of the somatic-marker hypothesis lies in its role in the decision making process. The somatic-marker hypothesis “serves to increase the accuracy and efficiency of decision-processes, screening out a range of alternatives and allowing the agent to choose from among fewer. ‘When a negative somatic marker is juxtaposed to a particular future outcome the combination functions as an alarm bell...’” (Bennett and Hacker, 2003: 212). This would suggest that a person who may have suffered a particular brain lesion or damage which may affect the somatic-marker process, would appear to be unable to distinguish between negative or positive felt-value of the mapped somatic-data and this affects her decision-making abilities.

Bennett and Hacker (2003: 213) disagree with the somatic-marker hypothesis by arguing that emotions are not caused by thoughts or somatic-data within the brain. According to Bennett and Hacker (2003:213), if the somatic-marker hypothesis is correct, then all terminology and diction used to describe emotion, such as anger and fear, would need to be re-evaluated in order to incorporate bodily states. It can be argued that no such thing is implied because ‘anger’, for example, might refer to a bodily state even though it does not seem to mean a bodily state.

Their argument is reminiscent of an old objection to physicalism: that it commits some sort of logical error. Place (1956) provides a compelling response to this objection in his essay Is Consciousness a Brain Process where he “argue[s] that
an acceptance of inner processes does not entail dualism and that the thesis that consciousness is a process in the brain cannot be dismissed on logical grounds” (Place, 1956: 55). He compares discussions about mind and brain processes to that of lightning and electric charges. Lightning and electric charges are one and the same thing but are referred to differently under different circumstances. So too are conceptual explanations of mental and brain processes, they are one and the same thing but used differently in different situations even though they both satisfy the same structure and function.

Again, considering Damasio’s work as an empirical analysis rather than a conceptual framework leads to Bennett and Hacker’s second disagreement. Bennett and Hacker (2003: 213) argue that if emotions are felt states caused by somatic-data, then one cannot be held accountable for the way one feels because being in a certain state and having a reason to be in that state are not concurrent. Bennett and Hacker (2003: 213) seem to suggest that Damasio claims one can be in an emotional state caused by somatic-data that has been mapped in the brain and for Bennett and Hacker (2003: 213), “given appropriate circumstances, we can say that someone ought to, and has good reason to, feel proud or ashamed, but we cannot say (save in what is merely a predictive sense), that his pulse rate ought to rise, or that his psychogalvanic reflect reactions ought to change”.


This counter-argument from Bennett and Hacker is not strong. Bennett and Hacker are equating states of somatic-data with certain physical states and this direct identity is not valid. Damasio is not claiming that all ‘feelings’ are accompanied by a certain physical reaction or that feelings should be undermined because they are physical in nature. Considering Damasio’s work as a conceptual framework, this kind of empirically-based rebuttal does not hold. Conceptually, the two ‘different’ states do not have to be interchangeable. Bennett and Hacker should not assume extensionality (i.e. substitutability of co-referring terms *salva veritate*) as it is not viable here. Somatic-data provides information about why a person should feel proud or ashamed; it is not actually the state of shame or pride. It is possible that ‘anger’ can be somatic-data, but this does not mean that anger means somatic-data. An example of this is to say that Parktown Prawns are terrifying, but it does not automatically mean that ‘to terrify’ means Parktown Prawns.

Other than interpreting emotion and feeling in a way that Damasio did not intend, Bennett and Hacker do not touch on the more philosophical issues regarding the idea of somatic-data and felt-value as being sufficient for consciousness or even consciousness-generation. I want to support Damasio here and grant the importance of somatic-data and felt-value as being paramount to consciousness-generation; however I still need to ask if this is enough to satisfy consciousness-
generation. It is, in theory, a novel way of showing how consciousness is brought about but it does very little to show how exactly we are aware of this process. How the protagonist interprets these feelings and understands them as meaning such and such is not clearly explained. For the purpose of this paper however, I am going to accept that somatic-data and felt-value are vital to consciousness-generation, but I am not willing to claim that it is entirely enough for consciousness.

8. Mental Maintenance

Many times philosophers contribute ideas about the mind without any real empirical evidence to substantiate their claims. Jesse Prinz (2012: 319) suggests that “perhaps if we begin by looking for an empirically motivated theory, we will end up with a better explanation of” consciousness. Prinz’s theory of consciousness rests on two components: first, that there is content to consciousness and an account thereof is possible. Secondly, content can become conscious. Where Damasio argues consciousness arises from different bodily-processes, Prinz argues that consciousness arises “only at an intermediate stage of perceptual processing” (Prinz, 2007: 185).

Similarly to other embodied cognitive theorists, Prinz explains that sensory systems are hierarchically ordered and that there is a succession of representations from “very local features of a stimulus” to “highly abstracted categorical
representations” and these highly abstract representations capture “features of stimuli that are comparatively invariant across perspectives” (Prinz, 2007: 185). Prinz suggests that there are representations that are neither local nor invariant, but rather coherent. It is here that he believes consciousness can be located. He claims that neural activity found in these intermediate-level areas “correlates with conscious experience, and cells in these areas have contents that agree with experience” (Prinz, 2007: 185). He claims that this intermediate-level hypothesis to be a theory of the content of consciousness.

I find this idea problematic. Stating that content of consciousness is developed at an intermediate-level is not sufficient in explaining what exactly this consciousness is. Also, using a term such as content implies, for me at least, some kind of *meaning*. Does it suffice to say: “this mug on my table has content” but never knowing that it contains coffee? Without giving meaning to content, we have books filled with letters and a world filled with objects. Providing content of something implies that one will know what the thing contains. If I am to give you the contents of this paper, I will tell you what it contains chapter-by-chapter and will not state simply that this paper is made up of letters. Or, this paper is where contents of a dissertation can be found.

Prinz shows us where content is generated and suggests that we become conscious of this content by “paying attention” (Prinz, 2007: 186). He mentions several
case-studies where subjects are asked to perform a certain task which requires a considerable amount of attention and haphazardly throughout the task they would ‘distract’ the subject in a very obvious way and surprisingly, the subject remained unaware of the distraction. “Putting this conclusion together with the conjecture about the content of consciousness, we are left with the following theory: phenomenal states are attended intermediate-level representations, or AIRs” (Prinz, 2007: 186). Prinz explains that attention is brought about through working memory and that working memory makes an allowance for manipulation and maintenance. This would mean that features we perceive become available for short periods of time and need somehow to be maintained and manipulated. Therefore, consciousness is generated when “perceptual states send signals to working memory” (Prinz, 2007: 187).

Prinz argues that being in a certain experience does not automatically result in having consciousness or the feeling of what it is like. He explains that working memory is the key to maintaining representations and that these representations are not “copies of sensory states” but rather are representations which are maintained or remain active even after the sensory perception has been removed. This kind of maintenance is possible through a process called working-memory encodings and that these encodings “forms the basis of phenomenal knowledge” (Prinz, 2012: 321). Finding an empirically motivated theory for consciousness
may still not answer some of the problems we have discussed here. I have already
discussed one issue with regards to contents of consciousness. Using
neurobiological terms to explain representations and suggesting that these
representations have content of sorts still brings us no closer in knowing what
consciousness is.

9. Where Neurobiology Fails

Traditionally, the mind is perceived as the faculty which allows an organism to
have a sense of self, a sense of consciousness, an ability to comprehend itself as
separate from its environment and also an organism which is able to understand
(even minimally so) its interactions with its environment. This mind is such that it
knows its own body in reference to objects outside of itself. And can also be
regarded as an experiencing and thinking thing. In addition, this mind is also
aware of a subjective character of experience. These minimal requirements for the
mind I discuss stems from a traditional view of mind which is very much
influenced by Cartesian thought.

Descartes (1637), in his *Meditations* attempted to establish fundamental principles
in which to approach all philosophy. His motivation was epistemological in nature
in that he desired to find the first knowledge of all things by putting science on a
sure footing. Essentially, being a mathematician himself, he endeavoured to
explain the knowledge of all things through philosophy just as science had
attempted to do before. He began his quest for attaining the fundamental principles through the conscious decision to doubt everything. This doubting process led to the realisation that he could not doubt the fact that he was thinking (doubting being a thinking process). Descartes established that “the mind by which I am what I am is wholly distinct from the body…and that although the latter [body] were not, it would still continue to be all that it is…” (Descartes, 1984: 44).

Descartes argued that the mind is able to exist without the body and that the mind is a thinking thing. This mind which Descartes discusses, is both eternal and immortal, and lives on after the body has expired. For Descartes (and other such dualists), the mind and the body are different in that the body is made up of a physical substance whereas the mind is made of a mental (non-physical) substance. One of the issues in the mind-body debate was introduced through this particular dualist theory because this separation leads us to ask how this non-material thing (mind) is able to interact with and have a causal relationship with the physical body. Descartes claimed that the mind interacts with the brain at the point of the pineal gland. Suggesting the pineal gland as the point of interaction still does not answer the question of how the non-physical mind is able to interact with the pineal gland (it being physical). Given the problem of interaction and
causation, physicalists have tried to provide their own set of answers to the mind-body problem.

Traditionally, and certainly stemming from a Cartesian view, being in possession of consciousness (from lower-order consciousness to higher-order consciousness) denotes a mind. This form of consciousness is such that we are able to place ourselves in the world, we are able to self-reflect and have a degree of self-awareness. Self awareness and the ability to self-reflect rely somewhat on the ability to think of oneself in a given situation. Within the Cartesian understanding, let this be the second claim for the concept of mind, that to have a mind we must be conscious and aware, and thinking is an important aspect of this. Moving from this supposition, we could assume that things which are not conscious or do not have an awareness of self also do not think.

In addition, the ability to be aware of others and other objects in relation to a self would depend on the thing possessing some form of a sense of self which would include the awareness or sense of subjective experience. Descartes’ viewed the mind as a thinking thing which had a unique view of itself, almost a knower of a self. The Cartesian way of thinking about mind is somewhat different from what proponents of embodied cognition consider the mind to be, or at least they do not share the same view with regards to how this mind is understood or brought about. Cartesian thought would hold that our minds are thinking things and
theorists of embodied cognition would claim that our minds are experiencing or feeling things.

The traditional view of the mind (stemming from Cartesian thought) is difficult to place within a working brain. Chalmers (2002: 247) claims this to be the central problem in the mind-body debate in that it seems that “[t]he central problem is that of locating mind with respect to the physical world”. Chalmers’ hard problems deal with those aspects of consciousness or experience that we cannot have third-person access to, nor can we understand other consciousness experiences as we understand own subjective perspectives. These problems include consciousness, subjective character of experience, phenomenal states and qualia.

Several philosophers of mind hold that mind is found or housed within the brain and some theorists (Smart, 1959; Place, 1956) argue that mental states are, or at least are related to, brain states. As we can see, the brain seems to be the main focus of the mind-body problem. Several kinds of physicalism look to the brain for answers regarding both the easy and hard-problems of consciousness and this is why I am looking to neuroscience and neurobiology to show us how it can contribute (however remotely) to these issues in philosophy of mind. I am not claiming that neurobiology will be able to provide an answer for these problems; nonetheless it is interesting to consider its contributions. These contributions may
prove to be very useful or completely useless regarding the harder problems of mind, but I think the concept of embodied cognition may offer some answers to philosophy of mind, or in some cases, it may introduce a new set of problems to philosophy of mind.

Some theorists in neuroscience and neurophilosophy argue for NCC and think that somehow finding neural correlates will answer the mind problem. However, we have established that finding the NCC will tell us that consciousness has neural correlates and that certain neuronal activity shows process of consciousness relations. This brings us no closer in answering exactly how this consciousness is brought about or even the purpose of consciousness in the human body. It shows us that damage to certain parts of the brain impairs certain consciousness episodes.

Further, some neurobiologists and neuroscientists have actively decided to move away from viewing the brain as the primary generator of consciousness and to study the body as being responsible for bringing about consciousness. Again, this may be getting us closer to understanding how consciousness is brought about but still tells us very little about the feeling process that Damasio and others proponents for embodied cognition have argued for. Also, suggesting that consciousness is generated from lower levels to higher levels is creative and plausible in theory, but what does this mean for questions concerning phenomenal
states and the feeling of what it is like? Stating that this is how consciousness is made does not tell us what it feels like or exactly what it means to even feel these things. It leaves me asking the question of how exactly this somatic-data is felt and subsequently understood?

In this chapter, I granted Damasio’s hypothesis that somatic-data and felt-value are key players in the construction of consciousness. I did not however readily agree that it is sufficient for consciousness-generation. However, if asked what I consider necessary conditions for consciousness, I could not provide an answer and thus have conceded that Damasio could be the best explanation from within the discipline of neurobiology at present. His answer may be sufficient for neurobiological purposes but it does not quite answer the accompanying philosophical problems regarding how we are able to feel these things and also how we interpret them into value-based meaning.

10. Conclusion

Neurobiology tries to contribute to the mind debate by suggesting that there are NCC, that consciousness has many different levels in the body, and that consciousness is an embodied process. In addition, neurobiology and neuroscience explain how consciousness is partly generated and also how it can be monitored through examining persons with brain lesions and other such damage. However, these kinds of explanations mean very little when we try to
understand consciousness as a sense of self or as the feeling of what it is like.

These explanations are lacking in some very important content: what exactly this feeling of self actually is. Rather, it provides us with empirical data that the body goes through certain processes when it experiences something, but it cannot tell us exactly what it is the body is in the process of experiencing.

It is easy to claim that something looks a certain way and even behaves a certain way, but rather difficult to say that this something is doing experiencing and also to give us sufficient content as to what it is experiencing. It is not possible (at least not yet) for neurobiology to tell me what experiencing is like, it cannot identify this feeling of what it is like, nor can it grasp it and manipulate it as it does other things. It is even debateable if neurobiology can tell me if consciousness exists somehow or somewhere given that if we are not ourselves conscious, we would not be able to know that there was consciousness, merely by studying neurobiology. Of course, we would not be able to study neurobiology if we were not conscious. This is problematic given that we need a certain property \( x \) in order to study this same property \( x \).

Suffice to say, neurobiology has offered a theory of embodied cognition which takes us one step closer to understanding how the body interacts with itself and its environment in a way which can be called aware or conscious. It provides us with a new understanding of consciousness as an embodied process rather than limiting
consciousness to one central part: the brain. From this chapter, an embodied
cognitive perspective has been established and discussed. This embodied
cognitive perspective is the way in which we view the world. It is necessary to
know ourselves and it is a hindrance in knowing others.
Chapter 2
What Experience Tells us about the Mind

What we learn through experience will tell us more about several pressing issues in philosophy of mind. Frank Jackson claims that we gain a new kind of knowledge (knowledge of non-physical facts, which I call non-physical knowledge for short) when we experience something, for example the colour red. David Lewis argues that a person learns no new physical facts upon experiencing red, but rather abilities such as the ability to recognize, imagine and remember. Jackson argues for the existence of non-physical knowledge and Lewis argues for the existence of acquired abilities. In this chapter I examine some of the counterarguments to Jackson’s Knowledge Argument and conclude that these philosophers have approached the Knowledge Argument incorrectly. I suggest a different physicalist account of experience.
1. What Experience Tells Us

We have established in the previous chapter that neurobiology contributes to the mind-body debate in a limited way. It may answer the easy problems of consciousness (such as reacting to environmental stimuli, the focus of attention etc.) but does very little to answer the hard problems of consciousness. That is, it does not tell us about how subjective experience is brought about or how we are aware of our sensory information. The inability to answer these harder problems of consciousness through such neurobiological enquiry has led some philosophers to argue that the mind eludes such scientific enquiry because it is not physical and therefore inaccessible to scientific exploration. Frank Jackson argues that the kind of knowledge we learn through experience tells us something about whether the mind is physical or non-physical. If we cannot access the mind the way we access other things, then we need to find another way to secure our knowledge of the anatomy of mind. The idea is if there are facts about the mind that are not among the physical facts, then the mind must have non-physical properties that those non-physical facts are about.

Jackson suggests we learn new knowledge or facts through experience and David Lewis seeks to refute this by suggesting that we learn certain abilities through experience. What we learn through experience may tell us more about how the mind functions and whether or not non-physical knowledge exists in addition to
physical knowledge. Until Lewis’ Ability Hypothesis has been refuted, Jackson’s argument for non-physical knowledge has not adequately been established. Lewis’ Ability Hypothesis occupies a logical gap that is left by Jackson’s Knowledge Argument. The fact that the Ability Hypothesis can be entertained illustrates that Jackson’s property dualism is not the only viable response or argument. If new knowledge is learnt upon experience and this knowledge is knowledge of non-physical facts, then Lewis’ account is false and Jackson is correct.

In this chapter, I will discuss Jackson’s Knowledge Argument and the most popular rebuttal to this argument: Lewis’ Ability Hypothesis. In this discussion, I will include other contributions to this debate and will conclude this chapter by showing that both Lewis and Jackson fail to address the problem of experience and what one learns through experience. Stemming from the discussion on neurobiological contributions in the first chapter and from this discussion on experience, I respond to Jackson’s Knowledge Argument in the third chapter by introducing a concept I call bifactualism, and offer a bifactual Account of Experience.

2. Frank Jackson: The Knowledge Argument

Jackson offered this particular thought-experiment, The Knowledge Argument, in *Epiphenomenal Qualia* in 1982. This argument was proposed as an argument for
dualism. Since then however, Jackson has retracted his dualist claim and has opted for a physicalist approach. For the purpose of this paper however, I would like to discuss Jackson’s original take on this debate.

Jackson tells us about Mary, the neuroscientist who has spent her life in a black and white room specialising in the neurophysiology of vision and is therefore an expert in wave-length combination of colours and how each colour stimulates the retina within the eye. After learning all the physical facts there are to know about colour and colour-perception, Mary is released from the black and white room. Here Jackson asks if it is possible that Mary’s knowledge of colour and colour-perception is actually complete even if she has never experienced seeing the colour herself before. The Knowledge Argument is therefore concerned “with the nature of Mary’s total body of knowledge before she is released: is it complete or do some facts escape it?” (Jackson, 1986: 278).

Jackson answers this question by proposing that Mary does indeed lack a certain kind of knowledge which cannot be gained by learning the physical facts of things. Jackson is saying that “physicalism is knowledge about the experiences of others, not about her own” (Jackson, 1986: 278). He goes further to state that if physicalism is true then Mary, by learning all the physical aspects of colour, knows all there is to know about colour. It is important to note the two claims this argument makes: Firstly, that before her release, Mary learnt and subsequently
knows everything there is to know about the physical nature of colour and colour-perception; secondly, that when Mary escapes from her room, upon experience, she learns something new. Therefore, what she learns cannot be a physical fact because she knows all the physical facts already. By extension, if Mary is able to learn non-physical facts, then it tells us that her mind must also at least have some non-physical aspects. Jackson’s Knowledge Argument is a strong argument for the dualist perspective and physicalists have struggled to establish counterarguments to weaken claims made through this thought-experiment. One of the strongest counterarguments is David Lewis’ Ability Hypothesis.

3. **David Lewis: The Ability Hypothesis**

David Lewis, in his article *What Experience Teaches* (1988) challenges the claim that non-physical facts are learnt through experience. In particular, Lewis questions the central claim that a fact is learnt when something is experienced. Lewis does not agree that a fact is necessarily learnt. He provides the reader with an example: he imagines that a concoction be made that smells exactly the same as a skunk. If a person were to smell this concoction, the neuronal activity would fire in the same way as if the person smelt a real skunk. Physically, her bodily processes of smelling the concoction would be the same as if she smelt a real skunk. This means that this person does not learn a new fact, no matter authentic the experience of smelling a skunk-like odour is. For Lewis, she would still have
the ability to recognise, imagine and remember what a skunk smells like regardless of whether or not she came into contact with an actual skunk.

Lewis admits that the experience of encountering a skunk and smelling it in the wilderness is somewhat different from smelling a skunk-like odour from a bottle. The two different spatiotemporal experiences aside, Lewis is suggesting that the neuronal activity resulting from the smelling process is the same. For Lewis, the physical occurrence behind learning something through experience does not have the same kind of results as learning in the black and white room, given that if one were to smell a skunk-like odour, it would teach you very little about the chemistry (therefore facts) behind the smell. If smelling a skunk-like odour means learning no new facts about its chemistry, then apparently no new facts are learnt. The kind of learning Mary does in the black and white room (classroom learning) is different from the content learnt outside of the room (experiential learning), but it remains debateable if Mary learns new facts upon experience. For Lewis, smelling a skunk does not tell you of facts about skunks, it could just mean you have now acquired the ability to recognise a skunk smell. Tye (2000: 226-227) states that Lewis “explicitly asserts that what experience E is like is the same as E. So, what it's like, according to Lewis, is a matter of fact”. Smelling the skunk-concoction is to have abilities: to remember, to imagine and to recognise the skunk-odour.
According to Lewis, by having a new experience, a person gains the ability to recognise, remember and imagine. This suggestion undermines the Knowledge Argument because the Knowledge Argument suggests that you learn new facts through experience. Lewis argues that there is a gap in Jackson’s argument because new facts are not learnt, but rather an ability to remember, imagine and even recognise objects as being related somehow to others. The ability gained for Lewis, is certainly not non-physical and gaining abilities is not the process of acquiring knowledge of new facts. This ability encompasses the egocentric proposition as well as the description of the experience itself. The egocentric proposition, according to Lewis (1988: 284) can be known only by a person seeing the colour while the colour is being seen. He gives an example of someone secretly slipping vegemite into his food and his having the ability to experience it, but not knowing that it is vegemite that he is experiencing. In this situation, for Lewis, no new non-physical information is gained but rather an ability to recognise, remember or imagine a strange new taste.

The vegemite example for me shows Lewis’ point better than the skunk example. I think Lewis is trying to prove that experience is only the best teacher when it comes to gaining abilities, but it is not necessarily the best teacher when wanting to learn facts. He does this by showing what we learn through experience may yield false results. The skunk example tries to show us that after sniffing the
concoction, we now do know what a skunk smells like without ever having to ever encounter a skunk. However, we learn no facts about skunk-like smells, just abilities relating to skunk-like smells. The vegemite example is perhaps more promising. It is suggesting that we may taste vegemite and never learn what it actually is, thereby not learn any facts about it. Thus our experience really only taught us what vegemite tastes like and gives us the ability to say in future (when we taste vegemite again): “I don’t know what it is I’m tasting but I recognise this taste, I remember this taste, and I can imagine it”.

According to Lewis, experience will not teach us new facts about things like vegemite, but it does teach us certain abilities. Lewis suggests that abilities do not mean we acquire any kind of knowledge of new facts, because gaining ability may mean we never know what fact it is we are experiencing, therefore no new fact is learnt. Lewis comments:

...knowing what it is like isn’t the possession of information at all. It isn’t the elimination of any hitherto open possibilities. Rather, knowing what it is like is the possession of abilities: abilities to recognise, abilities to imagine, abilities to predict one’s behaviour by imaginative experiments (Lewis, 1983b: 131).
4. Michael Tye: A Challenge to Both Jackson and Lewis

Lewis’ Ability Hypothesis provides a strong challenge to Jackson’s Knowledge Argument. It shows that we do not necessarily learn any new kind of fact through experience, since our learning might consist in developing abilities through experience. Michael Tye in *Knowing what it is Like: The ability Hypothesis and the Knowledge Argument* (2000) disputes Lewis’s Ability Hypothesis. He claims that Mary does not have the abilities that Lewis lists. When Mary first sees a red rose, she is able to say “this is what a red looks like” which is understood as an indexical concept. Tye argues that not only is she seeing just any shade of red, but a very particular hue called red17. She comes to know red17 as an indexical concept which means she only knows that particular shade of red while she is in the process of seeing red17.

Tye suggests that Mary can only know red17 as an indexical concept because she has no non-indexical conceptual knowledge of red 17 and this means she can only recognise red17 when she is in the process of seeing it. Knowing red17 as an indexical concept is such that if she should see two tomatoes in the same packet, one being red17 and the other red18, she would be able to pick out which one is red17 and which is red18 because she is experiencing these colours as indexical concepts. However, if a red17 tomato was presented to her on its own, she would
not have the ability to know if it were either red17 or red 18 without the ability to know both in relation to each other as indexical concepts (Tye, 2000: 230).

Tye concludes that Mary does not have the abilities Lewis establishes. He claims that Mary can only know red17 as an indexical concept and therefore does not learn the abilities to recognise, imagine and remember. Tye states that learning requires some kind of retention and Mary does not retain any knowledge of red17 and therefore does not learn a new ability. However, according to Tye (2000: 232) “the knowledge argument against physicalism is just that: an argument for knowledge. It makes no essential use of the concept of learning”. Tye is arguing that Mary does no learning of non-indexical concepts and therefore does not learn the abilities that Lewis claims because she cannot imagine, remember or recognise red17. If Tye is right, I imagine that Mary will then learn absolutely nothing upon her release and the only knowledge she has is what she learns in the black and white room. Incidentally this supports physicalism, contrary to what Tye claims.

Tye displays some awareness of this mistake in this particular article and executes something of a roundabout turn by proposing a revised version of the Ability Hypothesis. He claims that Mary does learn some abilities, just not all the one’s Lewis suggests. She gains the ability to recognise and the ability to cognize experiences as long as she is experiencing them. He suggests that Mary’s ability is
rather to apply an indexical concept to her experiences through introspection. He further claims that because this indexical concept can only be formed whilst Mary is in the process of having the experience, she obviously cannot claim to have the phenomenal concept of red in the black and white room. Therefore, Mary cannot ever think: “ah, this is what it is like to see red” in the black and white room because she cannot form this indexical concept. When she leaves the black and white room however, Mary makes a new discovery of seeing red and is then able to make the indexical claim. Therefore she is discovering something new, something she did not know in the black and white room.

I find Tye’s argument unsatisfying. He moves from claiming that Lewis is wrong because Mary cannot possibly learn abilities given that what she learns is purely indexical and lacks content of non-indexical concepts and therefore struggles to retain knowledge of an indexical sort. If Mary is unable to retain knowledge of the things she is experiencing, then she would not be a very successful neuroscientist given that even her classroom learning would certainly also contain indexical concepts such as: “this is the book I am reading from, this is what this machine is telling me” etc. Secondly, he then seems to retract this argument and proposes that she does learn abilities, but just not the ones Lewis suggests. She learns how to apply an indexical concept to her experience.
I do not quite know what this means with regards to knowledge she learns upon her release. If Mary learns a kind of know-how, which Tye seems to be arguing for, and if she undergoes a cognitive difference (which he also argues for), then something new is learnt – either fact or ability. If Mary undergoes a cognitive difference I would like to know what caused this cognitive difference: a new ability or new learnt-fact? Tye does not quite answer this. Stating that she acquires the ability to apply indexical concepts has me wondering how she thought of herself in the black and white room without indexical concepts.

5. Coleman: Why the Ability Hypothesis is Problematic

Sam Coleman (2009) writes in *Why the Ability Hypothesis is Best Forgotten* that the Ability Hypothesis does not offer a satisfactory rebuttal of the Knowledge Argument. Coleman argues that Lewis’ three abilities gained through experience are not necessarily adequate or can be combined. He suggests that to be able to remember something entails imagining it. Therefore, those two abilities can be combined to become one or we should choose one and discard the others.

The problem with the challenge Coleman presents here is that it is a weak attempt to break down the Ability Hypothesis. Lewis could have listed hundreds of abilities rather than the three he mentioned. Stating that Lewis mentioned too few abilities and that some can be combined does not weaken the argument that some abilities or other are learned rather than a new fact. The Knowledge Argument,
and by extension the Ability Hypothesis is concerned with the nature of what is learnt, not the volume of things learnt.

Coleman’s second objection to the Ability hypothesis is that “one might think what happens to Mary is that she gains abilities the acquisition of which necessarily involves learning a new fact- in which case physicalism would again be threatened” (Coleman, 2009: 4). My concern with Coleman’s objection is that he would need to clearly show that these abilities are not physical. The problem with Coleman, and with other such philosophers, is that he accepts Jackson’s first premise as correct and therefore concludes that if Mary learns abilities and somehow she has undergone some cognitive development through experience, then she has learnt a new fact.

If we accept Jackson’s first premise, then it is possible that learning something new over and above gaining ability would be a threat to physicalism. If one were to doubt Jackson’s first premise, then learning something new upon her release does not necessarily conclude that physicalism is threatened. Secondly, in order for new knowledge to be a threat, Knowledge Argument proponents would need to properly identify the new learnt fact as non-physical and not just another form of physical knowledge. It is also not enough to say that the knowledge acquired through experience is phenomenal as this does not validate either physicalist or dualist claims. Labelling experience as phenomenal means that the experience is
inaccessible to third person enquiry and is purely subjective and says very little about the physical or non-physical status thereof.

6. Brian Loar: Phenomenal States

Brian Loar (1990) in *Phenomenal States* discusses the prospect of consciousness and phenomenal qualities as being unreal because they are “irreducible non-physical-functional facts” (Loar, 1990: 296). He suggests that one should accept the irreducible nature of introspective phenomenal states as well as “take phenomenal qualities to be identical with physical-functional properties of the sort envisaged by contemporary brain science” (Loar, 1990: 296). Coleman also discusses the role of phenomenal states in learning how to ride a bicycle and asks whether or not an ability or new knowledge is acquired. He suggests that the body goes through several processes when learning how to ride a bicycle and this learning depends on “cognitively monitoring the quality of new phenomenal states” (Coleman, 2009: 83). This cognitive monitoring of phenomenal states can be illuminated by appealing to Damasio’s work and the discussion of the embodied cognitive perspective in the first chapter of this paper.

According to Coleman (2009: 83), when one learns to ride a bicycle, “one tries to ride, and falls off. One tries again and falls again. Quickly one learns that a certain feeling- in fact that of toppling- means being imminently in for a tumble and a cut arm. Conversely, keeping the bike, however fleetingly, within the acceptable
parameters of uprightness comes with its own distinctive new sensation: that of undisturbed equilibrium, of not being impinged upon by overpowering forces to either side” (Coleman, 2009: 84). Loar (1990:229) considers this kind of feeling as phenomenal and suggests that these phenomenal states are recognitional concepts which point out certain internal properties. He claims that these are physical-functional properties of the brain.

If all that is felt are internal stimuli which are neurobiological in nature, then we can conclude that, if phenomenal states are the subjective character of experience and thus the feeling of ‘toppling over’, then phenomenal states may be neurobiological in nature too. It could be the case that these states are neurobiological, but then why can neurobiology not tell me more about the feeling of what it is like? Claiming that phenomenal states are recognitional concepts and are physical-functional properties of the brain does not tell me about the feeling of what it is like or what exactly the nature is of the facts Mary learns. If Coleman is right however and these phenomenal states are physical-functional properties of the brain, then Mary does not learn all physical facts in the black and white room because she needs to learn about objects and facts through knowledge-by-acquaintance as Churchland proposes and which we will discuss in section 8 of this chapter. She would need to learn the feeling of ‘toppling over’ and can only do this through experience.
7. Benjamin Kuipers: Complex Information Processing

In attempt to establish the concept of phenomenal states as physically belonging to the embodied process, Benjamin Kuipers in *Drinking from the Firehose of Experience* (2008) discusses the possibility of computational theories accounting for the phenomenon of consciousness. He suggests that “we humans are vastly more complex in terms of the numbers and variety of processes taking place. Perhaps subjective experience just is the operation of a highly complex information-processing mechanism with very high information content as well as input and output” (Kuipers, 2008: 164).

Let us say that subjective experience is “the operation of a highly complex information-processing mechanism”, does this mean we can know the subjective character of experience by studying this highly complex processing mechanism? I think the problem lies in the access of subjective experience. It is our inability to understand what it feels like for somebody else that will not allow us to know what Mary experiences when she leaves the black and white room. Stating that subjective experience is a “highly complex information-processing mechanism” tells us very little about what it feels like for the person experiencing or what exactly it is she is experiencing. Because we struggle to access this kind of subjective content with current scientific enquiry, we feel inclined to consider it as non-physical somehow, as Jackson proposes.
However, information processing may be considered phenomenal, this does not explain how a person experiences the ‘toppling over’ feeling Coleman discusses previously. It is too simple to assign subjective experience to complex information-processing if you cannot explain what exactly this complex information-processing is. Labelling an instance or occurrence does not define the instance or occurrence. Mary could learn about this complex information-process objectively and still not know much about the feeling thereof. A label, as in this regard, fails to provide meaning or content and thus does not quite answer problems arising from the Knowledge Argument.

8. Churchland: Knowing-by-Acquaintance

The Knowledge Argument makes two important claims. Firstly, that before her release, Mary learns and subsequently knows everything there is to know about the physical facts of seeing red. Secondly, that when Mary escapes from her room, upon experience, she learns something new. From these premises, it follows that what she learns cannot be a physical fact because she knows all the physical facts already. The Knowledge Argument can only work if the first claim is granted as correct. Most philosophers focus their attention on the second claim and argue that either Mary learns new knowledge or does not learn new knowledge. Lewis is one such. He concerns himself with what exactly is learned upon her release and thus seems to also accept Jackson’s first claim.
Churchland’s account of the Knowledge Argument is such that I agree with the conclusion and not with the first premise. This argument at first glance seems sound. However, there almost seems to be some kind of trickery in the formulation of this argument. This argument rests on deriving a conclusion that follows from premises we assume are true. Generally, and especially in this kind of thought-experiment, the truth of the premises should not matter to the validity of the argument. However, I feel the need to argue that the premise allows no room for physicalism, regardless of the conclusion that is reached. The point is that although this argument proves to be logically sound, there must be something wrong with it, given that a similar argument establishes physicalism, as I am about to show.

Churchland (1989:2) claims that Jackson’s argument is formally valid, but that the argument “continues to see the same equivocation found in [earlier] castings of his argument”. Churchland suggests that Mary can have a complete “knowledge by description” in the black and white room but is missing knowledge by acquaintance. Churchland (1989: 4) claims that the “canonical presentation of the knowledge argument” would be just as valid if Jackson had claimed the following:

“(1) 'E' stood for 'is about something ectoplasmic in character' (where 'ectoplasm' is an arbitrary name for the dualist's nonphysical substance), and
(2) the story is altered so that Mary becomes an exhaustive expert on a completed *ectoplasmic* science of human nature”.

Having descriptive knowledge of this “nonphysical substance” would still not allow Mary any *know-by-acquaintance* “what it is like to see red”. She would have the same knowledge of what it is like to see red as she would if she had sat through discursive lectures on the physical, molecular facts of seeing red. Churchland calls this the “parallel knowledge argument” (Churchland, 1989: 4). It shows that having classroom or textbook knowledge of non-physical facts in the room would bring Mary no closer to *knowing what it is like to see red*. This is important for this discussion.

Churchland suggests that the argument is set out in such a way that it does not matter what Mary learns in the black and white room, there is something to learn outside of the black and white room, be it physical or non-physical. If she learns all physical facts in the black and white room, she is still lacking knowledge by acquaintance. Alternatively, if she learns all non-physical facts in the black and white room, she is again lacking knowledge by acquaintance. The problem here is not the kind of knowledge she learns in the black and white room, but the lack of knowledge by acquaintance which cannot happen in the room.

What this means for Jackson’s Knowledge Argument is that Mary is always lacking some kind of knowledge in the black and white room regardless of
whether or not she is learning physical facts or ectoplasmic science. This would mean that we must either reject the idea that Mary learns all things of a certain type in the black and white room and will learn something else outside of the black and white room or we must continue struggling to figure out what exactly she learns upon her release based on a strong premise that she learns all physical facts.

I take Churchland’s argument to show that there is something wrong with the first premise. I suggest that it should be re-evaluated, and instead of starting with the second premise, like so many do, I prefer here to dismantle this argument from the first. I am going to suggest that Mary’s knowledge of physical facts is not complete in the black and white room and will show how this is possible in a moment. This leaves me able to agree with Jackson that Mary learns something new upon her release. But I will argue, in the next chapter, that the knowledge gained is still knowledge of physical facts and thus premise 1 is false. There I will also offer an account of which physical facts Mary is excluded from learning in the black and white room, and why. In order to have prepared the ground for this argument, I have summarised existing theories of what Mary could have learnt when leaving the black and white room and now will discuss why the Knowledge Argument does not work.
9. Why the Knowledge Argument is Problematic

Lewis claims that the Knowledge Argument does not work because Mary learns abilities rather than any new kind of fact. Tye disagrees with both Jackson and Lewis, he claims that if Mary were to see the colour red17 she would not be able to remember it and that she only knows red17 as an indexical concept. She knows it while seeing it. This is closely related to Churchland’s notion of knowledge-by-acquaintance which proposes that Mary learns only descriptive knowledge in the black and white room and knowledge-by-acquaintance outside of the black and white room. Knowledge-by-acquaintance is to know something as an indexical concept. Loar and Coleman show how this knowing-by-acquaintance or knowing something as an indexical concept is possible by discussing the role of phenomenal states. Finally Kuipers attempts to prove that these phenomenal states are physical as an embodied process and for this reason, what she learns should be considered knowledge of physical facts.

So far we have discussed two different kinds of knowing that Mary can undergo. Mary can learn knowledge-by-description in the black and white room and she can learn knowledge-by-acquaintance once she experiences something. This debate is mainly centred on what Mary learns upon leaving the room. As mentioned before, this debate is fuelled by the fact that most theorists accept Jackson’s first premise as correct and this is what I would like to challenge. If we
revisit Jackson’s first premise, there may be a good chance that all the speculation about the physical or non-physical nature of what she learns no longer applies.

Prinz (2012: 294) states that the problem with the Knowledge Argument is that “it illicitly draws a metaphysical conclusion from an epistemological premise”. He also suggests that the Knowledge Argument is so debateable not based on its fallaciousness but rather that any responses to the Knowledge Argument appear problematic. Prinz argues that rejecting the first premise, as I wish to do is unpromising. He suggests that one could only reject the first premise based on the ambiguity of the meaning of ‘physical fact’. He states that if there are physical facts of the world that can only be known through phenomenological states, than those facts remain unknown to Mary in the black and white room. I agree with this.

He argues that dualists can counter this by suggesting that “let a fact be a state of affairs, and let a physical fact be any state of affairs that can be described in the language of the completed physical science, using no phenomenological vocabulary (the assumption made by standard physicalists is that phenomenological properties will ultimately be reduced to some other properties). Let’s presume that Mary knows all these facts. P1 is secure” (Prinz, 2012: 298). This would mean that Mary knows all kinds of facts which can be expressed in the language of physics. If Mary knows all these kinds of facts, then Prinz
suggests the first premise cannot be rejected and physicalists find themselves in trouble again. Prinz argues that all things that can be explained through physics vocabulary are physical and all things that cannot be explained in physics terms are non-physical. I am going to discuss this idea in more detail, and ultimately reject it, in the next chapter.

Prinz further identifies two kinds of knowledge that have been discussed:
Knowledge by description and knowledge by acquaintance. We have already discussed this with regards to Churchland’s contributions. However, Prinz states that knowledge by acquaintance “is a matter of having a special kind of epistemic access to that mode’. On this approach, Mary does not learn any new properties, but rather she acquires a new kind of knowing” (Prinz, 2012: 315). He also states that “when you shift from description to acquaintance, the only thing that must be added is some kind of encounter with the object of knowledge” (Prinz, 2012: 315). Prinz uses the idea of knowledge-by-acquaintance to prove that the first premise should not be rejected. He gives an example of providing a person with enough description of something that the change in knowledge through acquaintance is insubstantial. I am going to use this idea of “some kind of encounter with the object of knowledge” to support my rejection of the first premise and will show how I do this in the next chapter.
Prinz argues that philosophers have entered this debate from the “armchair” and therefore have not tried to show empirically what it is Mary learns through experience. He then expands on this by giving a neurobiological account of how Mary experiences something by looking at how her brain functions in the process of learning classroom-facts and then experiential facts. He shows that there is a physical difference with what occurs in Mary’s brain in the black and white room and outside of the black and white room. He claims this to be empirical evidence that Mary learns new things through experience. I agree that neuronal activity may show that she is learning something. Empirical evidence of neurobiological processes generally supports physicalism nicely but it is not a strong rebuttal to the Knowledge Argument.

The Knowledge Argument is about knowledge that Mary learns upon her release. Monitoring neuronal activity cannot be translated into the kind of facts that Mary learns. Firstly, stating that Mary learns something new upon her release does not tell us what it is she learns. Prinz attempts to provide us with some kind of empirical data of how Mary learns, but I claim this story still tells us very little about what she learns. Secondly, knowing how her brain reacts to certain experiential situations does very little in explaining what it feels like for her. It does not even touch sides with regards to the subjective character of experience and the feeling of what it is like. The subjective character of experience cannot be
discarded in this discussion. Let’s agree that what Mary experiences is: 1) knowledge-by-acquaintance, 2) is phenomenal in nature, 3) neuronal and therefore physical. What does this mean for the account of experience?

Jackson’s Knowledge Argument fails when a person rejects the first premise. The problem with rejecting the first premise by claiming that Mary does not learn all physical facts in the black and white room, is then to furnish the reader with an account of what physical facts she comes to learn upon her release. This is no easy task and by studying fMRI scans of the first time she sees red and saying: “Aha look! Her neurons are firing differently from when she was just learning classroom learning of seeing red, this means she must be learning something new!” is not enough to tell us exactly what she is learning. This is where physics explanation ends. It ends with showing that Mary’s neurons are firing differently and interpreting this as her learning something new. It can provide us with the empirical data Prinz advocates, but it can tell us very little of what it is like for Mary and what it is she learns.

In the first chapter of this paper, we discussed the role of neurobiology and showed that neurobiology cannot prove physicalism. In this chapter (two), we have discussed Jackson’s Knowledge Argument and have shown that this argument cannot disprove physicalism nor prove dualism. We need to ask ourselves if neurobiology fails because there are certain facts that escape it and if
the Knowledge Argument fails because it makes too strong a claim on what physicalists are able to know.

10. Conclusion

The Knowledge Argument claims that a new kind of knowledge is learnt through or with experience and that this new kind of knowledge is knowledge of non-physical facts. The Ability Hypothesis argues otherwise. The Ability Hypothesis argues that Mary does not learn any new facts upon leaving the room, she only learns new abilities: to remember, to imagine and to recognise. For Lewis, this does not constitute new knowledge at all. According to Jackson, Mary knows all physical facts already and that means if she learns something new, then it must be non-physical. Jackson is able to show this by setting up an argument which I claim is problematic.

Tye, Coleman, Kuipers and Loar try to show that Mary learns something phenomenal upon her release but that these phenomenal facts are about physical properties. Again, these philosophers are concerned with what it is that Mary learns upon her release based on the understanding that the first premise is correct. And again I must argue that stating that Mary learns something \textit{phenomenal} upon her release still tells me as much as Prinz stating that her neuronal activities behave differently through experiencing the colour red rather than learning about it from a text-book.
The aim of this chapter was to find out if what we learn through experience will tell us anything about the mind. If learning *know-how* can only be done through experience, then the content of the knowledge-how will tell us more about the mind. If the knowledge-how is knowledge about non-physical things and the mind is able to learn about this, then there are elements of the mind which are non-physical. However, if this know-how is closely associated with the way the body learns through its embodied cognitive perspective, then we are learning more about our physical bodies performing actions, perceiving external objects and interacting with an external environment. This know-how of ourselves interacting with the world is a physical process. The problem here should not be trying to establish the physicalist or dualist perspective of what this subjective experience is, but rather if facts generated from this subjective experience are accessible to current physics.

The Knowledge Argument is about the *knowledge* Mary learns upon her release. Prinz (2012: 316) is correct when he argues that we need a “substantive account of the special epistemic relation [and] we need to explain what our direct epistemic access to experiences consist in”. He gives a neurobiological account of this direct epistemic access but cannot tell us the particular facts generated from this experience. I am arguing that there is still something missing in this kind of account. So, let us agree that Mary learns something upon her release from the
black and white room, call it *know-how* or *knowledge-by-acquaintance* or what have you, as how this learning happens does not matter as much as what it is she learns. I am going to help myself to this statement and ask an important question: why can physics not know these kinds of facts generated through an experience?

In the next chapter I argue for a bifactual account of experience: bifactualism. The bifactualist argues for two kinds of physical facts: general and particular. It undercuts the Knowledge Argument by showing that physics explores the world from a ‘unifactualist’ viewpoint, where any facts that are expressible belong to one specific class: general physical facts. I argue that Mary learns this kind of fact in the black and white room and then learns another kind of physical fact once outside of the black and white room: particular physical facts. Through bifactualism I hope to show that physics (and thereby the physicalist account) should make room for physical facts that elude expression in the language of physics.

So far, in this paper I have shown how neurobiology cannot answer some of the more pressing issues in philosophy of mind. I also explored one of the most famous debates about experience by looking at Jackson’s Knowledge Argument. I have argued that most philosophers have approached Jackson’s Knowledge Argument by granting the first premise as correct and this has led to a lot of confusion as to what kind of knowledge Mary learns: physical or non-physical. I
claim that this argument is more complex than that, and I will now argue that there are certain facts that escape Mary because these kinds of facts cannot be expressed in the language of physics.
Chapter 3
Bifactualism: A Physicalist Account

The Knowledge Argument is an argument for dualism that claims that there are both physical and non-physical facts which can be learnt through experience. I reject the Knowledge Argument and instead provide a new physicalist response I call ‘bifactualism’. Bifactualism is a physicalist account essentially comprising two elements. First, it distinguishes between two kinds of facts: general and particular facts. The second element is the claim that this distinction explains facts about consciousness. I disagree with the dualist claims made in the Knowledge Argument and show that the Knowledge Argument neither supports dualist claims nor does it reject bifactualist claims. Most contributions made with regards to the Knowledge Argument focus on what Mary is able to learn once outside the black and white room. Bifactualism is interested in what she is able to learn in the black and white room which makes this a novel approach to the Knowledge Argument.
1. The Importance of Facts

In the previous chapter we discussed Jackson’s hypothesis regarding Mary, the neuroscientist, who finds herself trapped in a black and white room. Let us grant that she learns all of one kind of fact and upon her release learns another kind of fact. This debate is important in philosophy of mind given that what Mary learns in her black and white room and what she comes to learn upon her release will tell us something about the mind. The understanding is that what is learnt from or through a given experience will tell us more about the nature of the mind. Thus, a lot of time and effort has been vested in Frank Jackson’s Knowledge Argument and the kind of fact she learns.

In this chapter I introduce a view I call ‘bifactualism’. Bifactualism is a physicalist account which essentially comprises two elements. First, it distinguishes between two kinds of facts: general facts and particular facts. The second element of bifactualism is the claim that this distinction explains facts about consciousness. In the first section, I distinguish two different kinds of facts: general and particular. An example of a general fact is the fact that all ravens are black and an example of a particular fact is the fact that Ronny the Raven is black. Once I have explained the distinction between general and particular facts, I will then discuss how physical facts come in both general and particular varieties. I seek to establish in particular that it is not the case that all physical facts are
general facts. Once I have unpacked the different kinds of facts and established that they are both physical, I will then move onto the second section of this chapter: Applying bifactualism to consciousness.

In the second section, I give my own interpretation of Jackson’s Knowledge Argument and discuss the role of physicalism in his argument. I show how physicalism may rely too heavily (if not completely) on physics vocabulary to explain the physical world. I argue that physics, with its progressive change from medieval impetus physics to modern quantum physics does not allow us to assume a perfect account of facts about the world. It may be the case that the physicalist account makes claims based on a physics vocabulary and fails to accommodate other kinds of facts that physics cannot grasp. This is not because these facts somehow fall into a category of non-physical. Rather, physics is not in the business of categorising all facts in the world which is no failing of physics but is rather just the nature of the endeavour.

A) Bifactualism: A Physicalist Account

Bifactualism is a physicalist account proposing two kinds of facts which may account for conscious experience. The bifactualist asserts that there is not just one kind of fact in the world which the current physicalist may make claim to, but rather that there are two kinds of facts that can be known about the world. The two kinds of facts available are of a general and of a particular kind. This is not a
contentious view and most can accept this assertion that physics represents the world in generalities and it is possible that the particularity of objects in this world cannot be documented in physics. This would mean that a story told about an object, class or subject can have both general and particular facts. An example of this would be to claim that in general, whales are mammals, and Moby Dick in particular, is a mammal.

General facts are claims we make of the world by using quantified generalities. We make claims like “all apples are fruit”. This can be considered a general fact about apples and fruit. Or, “some fruit are apples” could be another example of referring to apples or fruit as generalities, this time using existential rather than universal quantification. Another example: no human being is immortal. This is a general fact about human beings. We apply generality to most objects and concepts in order to grasp a general understanding of the way the world is represented. General facts are those which can be expressed with universal quantifiers. Universal quantifiers are statements about all or none of a group. To show that a universal quantifier is true, we need to determine or verify that the statement is true for every value in the group. So, all A’s are B’s is a universally quantified statement. Universal quantifiers can be “read for all \( x \), if \( x \) is a philosopher then \( x \) is wise” or “every philosopher is wise” (Mautner, 2005: 634).
Existential quantifiers can be read as “some $x$ is $F$” or “there exists an $x$ such that $x$ is $F$”.

General facts are expressed in a general vocabulary using general classes and subjects. We know pens are tools used for writing because this is the generally understood and expressive use for a pen. We know tables exist in the world and they adhere to certain structures and functions because there are general facts about tables, their structures and functions. An example of a universally quantified statement is: all whales ($\alpha$) are mammals ($\beta$). An example of a statement which is not universally quantified is: Moby Dick (an instance of $\alpha$) is a mammal $\beta$. General facts cannot tell me that Moby Dick is a mammal. We say something like: all whales are mammals, Moby Dick is a whale, therefore he is a mammal. And the claim that “Moby Dick is a whale” is not a general fact, since it cannot be expressed without names or indexicals. General facts usually relate properties to individuals or objects. If this be the case, then we find that we still need to define Moby Dick as a general fact by claiming that he is a whale in order for the universally quantified statement to hold. We cannot say something like: all Moby Dicks are whales because there really is only one Moby Dick and therefore he cannot be expressed by general facts.

There are facts about Moby Dick which cannot be explained by using universally quantifiers. If we try to discuss facts about Moby Dick splashing about in the
ocean, we would need to say something like: whales swim in the ocean. This tells me that whales swim in the ocean and if I couple it with the claim that Moby Dick is a whale, I can deduce that he also swims in the ocean. But this still does not tell me what Moby Dick looks like as a particular, or where he swims in the ocean or when. The point of this exercise is not about how much detail the claim involves, but rather whether the claim is a general or particular.

It is possible that a particular claim can still lack detail and still not be expressible in the language of physics. Certain statements or claims cannot be expressed without a proper name or as an indexical and thus elude what is expressible in the language of physics. The movement from general to particular facts takes place by introducing statements of particular facts as premises. There are features of being a whale which can be described by using general facts. Features of the ocean can be described by using general facts. The relationship between a whale and the ocean can also be described in a general vocabulary. However, there is something particular about Moby Dick that general facts cannot capture.

Let us see if we can apply universally quantified statements to Moby Dick as a particular. Moby Dick is $MD$; The particular part of the ocean in which Moby Dick swims is $p$ (all this happens at time $t$). An attempted universally quantified statement could look something like this: all $MD$s can be found in $p$ at $t$. This universally quantified statement is tricky. $MD$ is referring to a particular whale
and physics is not in the business of categorising all the world’s whales and their respective names. So let us assume that Physics would then refer to Moby Dick as a whale and convert the above statement to: all whales can be found in \( p \) at \( t \). It is not possible that *all whales* can be found at a particular part of the ocean at a particular time. Physics could then change its tactic and claim that *some whales* can be found in \( p \) at \( t \). It is more possible that some whales can be found in \( p \) at \( t \). However, is Moby Dick one of them? Perhaps it is the case the Moby dick is one of the whales and perhaps he is not one of the whales found in \( p \) at \( t \).

Physics cannot tell us this because it uses general facts to explain the world. There are particular facts about Moby Dick that physics cannot tell us about. For every general fact about properties, a particular fact exists. Particular facts are about objects which exist at a particular place and time. These facts are particular or specific and therefore do not satisfy conditions of universally quantified statements. Universally quantified statements will make claims about all or some things, but never about one thing. There are particular facts about a particular object at a particular space and time which physics cannot tell me about. Let us look at the above example again. Moby Dick is a whale; Moby Dick is \( MD \); the particular part of the ocean in which Moby Dick swims is \( p \). All this happens at time \( t \). All a universally quantified statement can tell me as is that all whales are mammals: all \( x \)'s are \( y \)'s and that all \( x \)'s swim in \( p \)'s.
Referring to a particular would look something like this: \( MD \) swims in \( p \) at \( t \). In order to refer to a particular object, certain conditions need to be met: an object must have been instantiated, currently be instantiated or will be instantiated at a particular time and place. If we are unsure of the time, place or the property in question, then referring to the particular would be more difficult. An example of this would be to say that Moby Dick swam in the ocean in the 1800’s. This still refers to Moby Dick as a particular whale but leaves the question unanswered as to when he swam in the ocean and this results in an incomplete account of the particular. If we cannot know \textit{when} Moby Dick swam at that particular place, then are we sure he was there to begin with? There are particular facts about Moby Dick because we know him to be a whale and we know where to find him and when.

I am going to make use of another example to show the distinction between general and particular facts. All domestic cats are mammals. All cats are also known as felids or felines. Most cats are carnivorous and weigh between 2.5 – 9kg. All the facts mentioned above are universally quantified statements and all these facts are general. However, these general facts tell me very little about my own cat: Angie. Angie can be explained by using universally quantified statements because she satisfies the condition of being a cat and therefore can be explained in generalities. But there is more to know about Angie than what these
facts can tell me. Angie is on my couch at this particular time and place and therefore there are facts about Angie that are not general, but particular.

Bifactualism supports the notion that there are two kinds of facts for objects. General facts can be fully expressed with quantification. Particular facts are those which cannot be expressed with quantification, predicates, variables and logical connectives alone, but also require names or indexical expressions. My next move is to show that both these facts have physical instances. In other words, there are particular physical facts. It needs to be explained that it is not the case that all physical facts are general, but rather that it is possible that some physical facts are particular. Establishing this will help the reader understand the claims I make in the second section of this chapter.

In this section I have established that there are two kinds of facts: general and particular. I claim that particular facts cannot be expressed in the language of physics. In the next section I will argue that there are particular physical facts. From this, I conclude that there are physical facts which cannot be expressed in the language of physics.

2. General and Particular Facts are Physical

There is no single universally accepted account of what is considered physical. Most theorists look to physics to explain the physical nature of the world.
Physicalists for one, exhibit a certain commitment to physics and what can be expressed in the language of physics. Physicalists are committed to the view that all things are physical, are reducible to the physical or emerge from the physical. For the physicalist, nothing exists over and above the physical. Making this kind of claim can be tricky if establishing physicality of objects is proving to be problematic. In the following discussion, I show how establishing physicality is far from being complete and this is an area physicalism needs to consider redefining and establishing.

In this section, I am particularly interested in arguing for the physicality of facts. I do this because bifactualism is essentially a new physicalist account. However, when I write about traditional physicalist claims, I am inadvertently writing about claims made about or by physics. The physicalist view is committed to physics and its findings. If physics cannot provide certain information about certain properties, states or events then the physicalist claims that these properties, states or events do not exist. The understanding is that whatever can be eliminated by, or cannot be accounted for in physics does not exist. Physics cannot account for particular facts and thus some would argue that particular facts do not exist.

The relationship between the physicalist and the scientist is a complicated one and I would like to use an analogy to explain it. A neuroscientist uses an MRI scanner to tell her more about the brain. The scans that come from this scanner provide
her with information to complete her theories or to argue her theses. If, unbeknownst to the neuroscientist, the MRI scanner began to produce faulty scans, she would begin receiving incorrect information which would greatly affect the outcome of her work. After the first scan, she sees a shaded area which she assumes is some kind of lesion; she begins writing papers based on her findings. If after the 5th scan she does, she sees the same lesion reappearing over and over again, she may realise that the scanner is faulty. She should have the scanner repaired or make use of another.

This is akin to the relationship between the physicalist and physics. The 
Physicalist is the neuroscientist and physics is the scanner. Physics may produce incorrect or incomplete information at times and it is up to the physicalist to review what information is being provided, perhaps tweak it a bit, or replace it with another. It is perhaps also advisable for the physicalist to identify what exactly she considers as physical in this world. She could rely on physics to tell her this, but based on the progression of physics (taking into account both failure and success), I would say that physics should not be relied on blindly to provide an account or comprehensive answer.

Stoljar (2009: n.p) points out that there have been quite a few physical theories over time: medieval impetus physics, Cartesian Contact Mechanics, Newtonian physics and current quantum physics and that there are shared or overlapping
factors. Perhaps a physicalist view should be based on a strategy of selection, similar to what is being undertaken by certain theorists in scientific realism (Harker, 2013). In order to do this, the physical theories mentioned above would need to be dissected in order to identify each part or function that was responsible for the success of the theory at the time and also parts that were responsible for the progression into the succeeding theory. Once this has been accomplished, the theory could only be considered complete and perfect as long as physics makes no further progression. We know that physics is always progressing as the form and ability of enquiry does and therefore we can hold that the physicalist view is always in a state of modification but never complete. What I would like to see is a physicalist account which is able to make claims such as: “this is physical because it meets the following requirements or satisfies the following conditions”.

To show you that we do not have a universally acceptable way of establishing physicality of properties, Stoljar (2009: n.p) claims that there are two kinds of questions that arise from trying to explain claims made by the physicalist that all things are physical: 1) what does it mean when physicalism claims that everything is physical and 2) What conditions must be met for a property to be considered physical? Stoljar (2009: n.p) identifies two kinds of answers to what he calls the condition-question. The first kind of answer “ties the notion of a physical property to a notion of a physical theory” and Stoljar proceeds to call this the theory-based
conception of a physical property. This theory-based conception claims that “a property is physical iff it either is the sort of property that physical theory tells us about or else is a property which metaphysically (or logically) supervenes on the sort of property that physical theory tells us about” (Stoljar, 2009: n.p). Stoljar gives an example of this by suggesting that if physical theory can tell us about the properties of mass, then anything “having mass is a physical property” (Stoljar, 2009: n.p).

There are two problems with the above theory-based conception. This claim is problematic because it does not support the assertion that everything is physical. Implementing Churchland’s strategy, we can reverse the statement. A property is non-physical iff it either is the sort of property non-physical theory tells us about or else is a property which logically supervenes on the sort of property that non-physical theory tells us about. This particular formulation of the statement supports the existence of non-physical facts. To claim that something exists because a theory informs us of its existence is flawed. Universally quantified statements are not existential statements. An example is to provide a convincing theory of unicorns and then claim that unicorns exist because the theory claims so. This does not substantiate the existence of unicorns.

The second problem with the above theory-based conception is that establishing an account of physicality based on a physical theory is troublesome if the content
used to make up the theory is what is being questioned. I cannot see this working because it is similar to making the claim that a table is physical if it can be explained through a table-theory, incidentally this table-theory is dependent on the physicality of tables. Another example is to state that a theory of unicorns is dependent on unicorn-like statements, assuming the existence of unicorns. Physics gives an account of what is physical based on its theory and if the theory does not accommodate for certain facts, they are discarded. Thus, if the theory-based conception is correct, then physics has no terms for particulars, e.g. Moby Dick. According to the theory-based conception, if the theory cannot tell me about the particulars, then it follows that the particulars do not exist or are not physical. Therefore, the theory-based conception is not air tight.

The second kind of answer Stoljar discusses is more promising than the first. This answer “ties the notion of physical property to the notion of a physical object” (Stoljar, 2009: n.p). Stoljar calls this notion the “object-based conception”. The object-based conception holds that:

A property is physical iff: it either is the sort of property required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents or else is a property which metaphysically (or logically) supervenes on the sort of property required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents (Stoljar, 2009: n.p).
Stoljar gives an example of this by suggesting that if objects such as trees, rocks etc are paradigmatic physical objects, then properties of these objects are physical properties. Also, if the property of “having mass” is needed to explain the intrinsic nature of objects that are physical, then by extension having mass is also a physical property. I am concerned regarding the circularity of both these conceptions. By describing a table as an object which has table-like properties certainly gets us no closer to the definition of a table. Stoljar argues that these conceptions can only be considered circular if “the conceptions are interpreted as providing a reductive analysis of the notion of physical” (Stoljar, 2009: n.p).

Stoljar argues that providing an understanding of the physical should be sufficient without requiring a conceptual analysis.

This response is problematic because the physicalist mostly relies on reductive techniques to decide whether a property is physical (especially with regards to reducing mental properties) and yet these physical properties cannot undergo a reductive analysis themselves. Stoljar (2009: n.p) notes that “when philosophers attempt to provide an analysis of some concept or notion, they usually try to provide a reductive analysis of the notion in question, i.e. to analyse it in other terms”. This may have two potential outcomes: 1) the physicalist must find a way of applying its analytic techniques to itself and this would mean defining physical properties without using the term “physical” in its definition or 2) the physicalist
must redefine its analytic technique in order to give a better definition of a physical property.

Stoljar (2009: n.p) further discusses another problem with the above conceptions which he identifies as Hempel’s Dilemma. According to Stoljar, Carl Hempel “provided a classic formulation of this problem: if physicalism is defined via reference to contemporary physics, then it is false – after all, who thinks that contemporary physics is complete?” (Stoljar, 2009: n.p). Further, if physicalism is defined using a future, ideal physics, then it is vacuously true because we have no way of predicting the development of physics in the future. For Hempel, we have no clear definition of the physical on which to base our understanding and application. Stoljar’s response to this is to suggest that at least we have some form of physics to work with and that in some respects, certain parts of physics can be considered complete.

Another response to Hempel which Stoljar identifies is “that a particular proposal about how to define a physical property – namely, via reference to physics at a particular stage of its development—is wrong” (Stoljar, 2009: n.p). Stoljar argues that even though we may not have a particular proposal, we still do have some kind of notion of the physical. My concern here is that having some kind of notion of the physical is quite a weak statement and I wonder how physicalists are able to hold such a strong claim over dualists given that their view is primarily
committed to an incomplete physics and what it shows. Also, suggesting that referring to physics at a certain stage of development is wrong is interesting but is not going to help us decide what stage of development is successful or unsuccessful to claim this to be a wrong undertaking.

A good science is one which has expectations of failure and success and the project of mapping out the physical universe as such will encounter both failures and successes. Physics is a science which has expectations of failure and success and therefore is always in a state of progression. Physics expresses its findings using general facts; however there are some facts which are *not expressible in the language of physics* which are facts about the particular. It is not the case that “only propositions expressible in the language of physics can be true, or that any individuals in the domain of physics can exist” (Haugeland, 1998: 101). It is entirely possible that facts can exist outside of what is expressible in the language of physics.

It would be strange to claim that because physics cannot express that Moby Dick swam in the ocean off the Chilean island of Mocha at 15h00 on the 31st of July, 1815, that Moby Dick never existed merely for this reason (better reason, of course, would be that Moby Dick is a fictional character). It would be strange to claim that I do not exist because I am not expressible in the language of physics. The language of physics can tell me about the physical properties of the books
you have on your shelf, but not that they exist and are on your shelf. Similarly physics cannot tell me anything *particular* about the office in which you find yourself, or the chair in which you sit, or even *about you* reading this paper. It would be unwise of me to doubt your existence because physics can only offer a description of your physical properties and cannot express the particular facts about you, namely that you satisfy this description. The particular facts about you sitting in your office, reading this paper are surely as physical as the general facts about human beings sitting in offices reading papers.

Let us use Churchland’s strategy here and present the argument for particularities in reverse. If particular physical facts are non-physical because they cannot be expressed in the language of physics, then it follows that the fact that you hold that particular pen in your hand is a non-physical fact. The book on your shelf, because it cannot be expressed by the language of physics, is also a non-physical fact. Yet, all the physical properties of that state of affairs can be described by physics— a physical description can fully describe it, except for the particularity. This leaves us with a bizarre and paradoxical conclusion that a non-physical fact satisfies an exhaustive description in the language of physics. This obviously does not hold. It surely is the case that particular facts are physical facts.

Physics is not in the business of categorising all the objects in the world but rather its function is to show underlying patterns of objects in the world. Physics does
not make existential claims about objects in the world. It does not state that an object exists at a particular time and place but rather that objects having certain properties will behave in a certain way, thereby documenting laws of nature.

The Mill-Ramsey-Lewis account “proposes that laws of nature be regarded as axioms or theorems that appear in those deductive systems that strike the best balance between strength of description and simplicity” (Jackson and Smith, 2005: 797). The strength of a system is normally decided based on the amount of information it contains about the world. This would mean that the more information physics has about the world, the stronger it is as a deductive system.

However, “simplicity is measured by how efficiently the system organises the disparate facts that describe the universe” (Jackson and Smith, 2005: 797). If physics documented or catalogued all objects in the world, it would be a very strong deductive system but would lack almost entirely in simplicity and thus efficiency, rendering it quite useless. As things stand, physics concerns itself with documenting the general underlying patterns of all things, thus ensuring a good balance between simplicity and strength. In order to maintain this balance between simplicity and strength, physics cannot account for all the facts in the world. It can account for what can be quantified over and thus for general physical facts (to ensure the correct amount of simplicity) but it cannot account
for what cannot be quantified over and thus for particular physical facts (to ensure the correct amount of strength).

In the first section of this chapter I show how there are two kinds of facts that can be known about objects in the world: general physical facts and particular physical facts. General physical facts are known through universal quantifiers and are expressible in the language of physics. Particular physical facts are not expressible in the language of physics. However it is not the case that because these facts are not expressible in the language of physics that they do not exist or are somehow non-physical. There are particular (as well as general) facts and there are particular (as well as general) physical facts.

The bifactualist is making the claim that certain facts cannot be reduced to what can be expressed in the language of physics. Thus, it may seem that bifactualism is a form of non-reductive physicalism. Non-reductive physicalism holds that not all things are reducible to the physical and also that not all things can be expressed in the language of physics. Non-reductive physicalism maintains that all mental properties may not be reducible to the physical but they at least supervene on the physical. Lewis (1994) explains supervenience by asking us to:

Imagine a grid of a million tiny spots—pixels—each of which can be made light or dark. When some are light and some are dark, they form a picture, replete with interesting gestalt properties. The case evokes reductionist comments. Yes, the
picture really does exist. Yes, it really does have those gestalt properties. However the picture and properties reduce to the arrangement of the light and dark pixels. They are nothing over and above the pixels. They make nothing true that is not made true already by the pixels. They could go unmentioned in an inventory of what there is without thereby rendering that inventory incomplete. And so on (Lewis, 1994: 413-414).

Horgan (2001: 218) explains that the “picture reduces to the pixels” and the reason this happens is because the picture “supervenes on the pixels”. Lewis maintains that there would not be any change in the picture and its properties unless a change occurred in the shading of pixels. Horgan (2001: 218) argues that this kind of supervenience is a form of reductive physicalism. Horgan (2001: 218) defines reductive physicalism (and later non-reductive physicalism) as:

The view that i) all natural properties are physically natural properties, hence ii) there are no higher-order, multiply realizable, natural properties; and moreover iii) mental-state names are kind restricted non-rigid designators of physically natural properties. Let non-reductive [physicalism] be the view that i) not all natural properties are physically natural properties; ii) there are higher-order, multiply realizable, natural properties; and iii) mental-state names are rigid designators of certain higher-order, strongly multiply realizable, natural properties.
Horgan (2001: 218) explains that in this version of non-reductive physicalism there are some properties which should be treated as natural properties “in addition to the physically natural properties (that is as natural properties over and above the physically natural properties)”.

From the above information about non-reductive physicalism, I would like to focus on two aspects: i) there are certain facts that cannot be reducible to what can be expressed in the language of physics; and ii) that there are “natural properties which exist over and above physically natural properties” (Horgan, 2001: 218).

The non-reductive physicalist typically supposes that mental properties supervene on the physical which means that there is some relation between them such that different arrangements of mental properties require different arrangement of physical properties. The bifactualist is not denying this claim but is rather pointing out that there is another category of physical fact that does not reduce to the language of physics- not some special class of ‘mental properties’ as such, but rather particular physical facts. The bifactualist will further claim that facts about consciousness are amongst these, and that this explains why facts about consciousness cannot be expressed in the language of physics. It is thus similar to in some ways (perhaps even compatible) with non-reductive physicalism. However, the bifactualist is offering a different explanation of why the language of physics is inadequate to express mental facts.
B) Applying Bifactualism to Consciousness

3. Physical Facts in Jackson’s Knowledge Argument

Jackson’s Knowledge Argument has been interpreted again and again over the years and here I offer another interpretation based on his first account of the Knowledge Argument and not based on the response he offers to Churchland’s Three Objections. I choose to mention Jackson’s response to Churchland in a footnote so as to not detract from the purpose of this paper. Following, I am going to present two renditions of Jackson’s Knowledge Argument. I choose to present them as two versions because each has ramifications for the dualist account and possibly the current physicalist account.

---

10 As noted in the previous chapter, Jackson (1998) later retracted his dualist perspective regarding the Knowledge Argument and adopted a physicalist perspective: a priori physicalism. A priori physicalism holds that “mental truths are a priori entailed by physical truths” (Stoljar, 2001: 311). Jackson has, since then, written several papers on his new physicalist view (1994).

11 Jackson (1986: 279) gives the following canonical representation of the argument in response to Churchland’s Three Objections:

(1) Mary (before her release) knows everything physical there is to know about other people.

(2) Mary (before her release) does not know everything there is to know about other people (because she learns something about them on her release)

Therefore,

(3) There are truths about other people (and herself) which escapes the physicalist story.
Version A

| P1) Mary learns all physical facts in the black and white room. |

Version B

| P1') Mary learns all physical facts expressible in the language of physics in the black and white room. |

P2) Upon her release, Mary learns some new information.

Therefore

P3) Mary learns non-physical facts

Version A claims that Mary is able to learn all physical facts in the black and white room by reading about physical facts in books, watching monitors and perhaps even interpreting software available to her in the room. According to bifactualism, Mary cannot learn all physical facts in the black and white room as there are facts that cannot be expressed through the information available to her. Mary learns general physical facts in the black and white room and upon leaving the black and white room would have the opportunity to learn particular physical facts. Therefore, P1 should be rejected on the grounds that Mary could not possibly learn all physical facts in the black and white room. With regards to
version A, I reject P1 because Mary cannot learn all physical facts in the black and white room as she only learns general physical facts and particular physical facts remain unaccounted for. I further reject P3 based on the rejection of P1.

Version B however is trickier than version A. Jackson begins his paper by explaining that “it is undeniable that the physical, chemical and biological sciences have provided a great deal of information about the world we live in and about ourselves. [...] I will use the label ‘physical information’ for this kind of information, and also for information that automatically comes along with it” (1982: 273). Jackson continues by claiming that “there are certain features of the bodily sensations especially, but also of certain perceptual experiences, which no amount of purely physical information includes” (Jackson, 1982: 273).

Jackson then claims that “we have all the physical information. Therefore, knowing all this is not knowing everything about [Mary]. It follows that Physicalism leaves something out [and] physicalism is incomplete” (1982: 274). I interpret this as meaning: physicalism is committed to findings made by physics and if physics cannot provide all the necessary information of Mary’s experience of red, then physicalism is somehow flawed, or worse still, false. Mary learns all this kind of physical information in the black and white room and she comes to learn something new upon her release. Therefore, according to Jackson, the physicalist account is false.
Version B states that Mary learns all physical information expressible in the language of physics in the black and white room. For Jackson, this physical information includes information which physical, chemical and biological sciences have provided. He also states that Mary learns all “information that automatically comes along with it” (1982: 273). This is a strong claim because it seems to suggest that Mary learns all information that physics is able to provide now and in the future, thus a kind of perfect, complete physics; This kind of claim will look something like: the perfect science\(^{12}\) is one that can provide all physical information about the world and this is what Mary learns. From this strong premise, whatever Mary learns over and above this is non-physical.

From a physicalist perspective, P1\(^{1}\) of version B is a tough premise to challenge. Several philosophers (Lewis, Churchland, Dennett etc) are inclined to either challenge P2 or P3, but seldom do they challenge P1 or P1\(^{1}\). Prinz claims that “almost everyone agrees that premise 1 is beyond reproach” (Prinz, 2012: 297). He further claims that the only dispute that can possibly be put forward

\(^{12}\)The physicalist is committed to claims made by physics and not necessarily the sciences. Jackson mentions above. Therefore, to claim that all physical information provided by all these fields is what a physicalist is committed to is not entirely true. From here on out I will remain true to the physicalist account and refer to all physical information as information generated from or provided by physics. Also, it is understood that all sciences could, in theory, be reduced to physics.
regarding premise 1 is the ambiguous nature of physical facts. As already discussed in the previous section of this chapter, there are two kinds of physical facts and assuming that physics is able to provide information about all facts is wrong. Physics cannot account for particular physical facts.

According to bifactualism, P1' is correct and P3 does not following from P1' or P2 in version B. I am first going to discuss the problem with assuming a perfect physics and then I am going to explain why physics cannot provide all physical information of the world, thereby showing that Mary did not learn all physical information in the black and white room. Jackson assumes a perfect physics which is complete and thus is able to provide all the physical facts about the world currently and about the world in the future. I claim that perfect physics is one which has an expectation of failure and success. Jackson would claim that a perfect physics is one which expects nothing because it imagines itself to be complete. I argue that a physics which claims to be complete and expects no future successes or failures is no longer in the business of discovery and can therefore no longer be labelled science, rather it becomes a historical enquiry of sorts. Also, however perfect this physics will ever be, it will still not be able to express particular physical facts.

As discussed in the previous section, physics explains and defines the world by using general physical facts. A general physical fact is expressible in the language
of physics. Physics can tell me about physical facts by using universally quantified statements. However, physics cannot tell me much (if anything) about particular physical facts. In agreement with Jackson, there are facts about objects which escape the physicalist story. However in Jackson’s argument he is assuming a perfect physics which has no physical facts escape it and I am assuming a physics which is always in a state of progression and also unable to express facts about the particular.

With regards to version B, I accept P1' because it is possible that Mary learnt all physical facts expressible in the language of physics and I reject the conclusion because it does not follow that any new facts she learns are non-physical. She could learn particular physical facts, which physics could not tell her in the black and white room. I accept P1' based on the notion that particular facts escape what is expressible in the language of physics. But then P3 does not follow because learning non-physical facts does not follow from the acceptance of P1'.

For bifactualism, physics does not account for all physical facts in the world because it defines the world in generalities and by using quantified statements. I have shown that there is more to know in this world than this account. For Jackson, physics does account for all physical facts in the world. Let it suffice to say that either P1 is false, or P1' is correct, and that the conclusion is incompatible
with the bifactualist account. Therefore, the Knowledge Argument does not enable dualism, if anything it supports bifactualism.

Importantly, Jackson (1982) makes two claims at different times in his paper: “physicalism is false” and the “physicalist story is false”. Jackson is right in one regard and wrong in the other. The physicalist story is false because it relies on what can be expressed in the language of physics and we have already established that not all physical facts can be expressed in the language of physics. However, to claim that physicalism is false is different. The theory that all objects and aspects of this world are physical could very well be correct but scientific vocabulary is not comprehensible. Because a vocabulary may be unable to express certain physical facts does not mean the physicality of the fact is questionable. This is then related to the P1 where Jackson claims that Mary has access to all the physical facts in the room and I claim that she does not. There are certain facts that escape her because they are not expressible in the language of physics and these particular physical facts are what make this general vocabulary unreliable.

In this section, we have discussed two versions of Frank Jackson’s Knowledge Argument. I argue that P1 of version A should be rejected on the grounds that Mary cannot learn all physical facts in the black and white room because she only learns general physical facts. I further reject P3 of version A based on my
rejection of P1. Furthermore I argue that P1’ of version B should be accepted because it is possible that Mary learns all facts expressible in the language of physics, but then I argue that P3 of version B does not follow from P1’ and should therefore be rejected based on the grounds that there are facts that elude what is expressible in the language of physics. These facts are particular physical facts.

3.1 Bifactualism and Learning of Facts

General physical facts are what physics is capable of expressing because they can be described through quantified statements. General physical facts are facts about physical properties or events that are observable through physical enquiry and which can be explained with the use of physics vocabulary. These physical properties comfortably adhere to and obey laws of physics and can be understood to have relations with other properties and can be instantiated by, or cause a physical state/event or events. General physical facts are about what Lewis (1986: 60) calls ‘fundamental physical properties’. Lewis suggests that “physics has its short list of ‘fundamental physical properties’: the charges and masses of particles, also their so-called ‘spins’ and ‘colours’ and ‘flavours’, and maybe a few more have yet to be discovered” (Lewis, 1986: 60).

These general physical facts concerning these properties capture the anatomical structure and function of most things (objects, concepts etc) and can be stated in the language of physics. These kinds of physical facts are accountable and can be
explained in physics and they describe the general molecular structure of most
things in a kind of general time and space. The moon can be used as an example.
For human beings, when we look up at the night sky we see a suspended kind of
round, glowing thing that changes shape and size throughout the month. We know
this moon as we see it. This is not the way physics would know the moon.

Physics would refer to the moon by way of statistics, its movements, its
suspension, its gravitational effects etc. It would discuss the lunar material (95%
balsatic liquid) as well as any kind of samples brought back to earth from several
space-missions (Gast, 1971: 121). However, these facts will tell us very little
about the particular physics facts about the moon. It will tell me very little about
the moon at a particular time and space. Physics offers generalities and does not
provide for particularities of an object. It can tell us about pens, but it can tell us
very little about the pen you currently hold in your hand. It cannot tell me about
the particular pen in your particular hand at this exact time. It can tell me that
pens are devices used for writing and that they are approximately 8cm in length, 2
cm diameter (given the general pen is tube-like and not square or any other un-
pen-like shape) and has a rubber grip for easy writing.

Mary, in her black and white room picks up a text book and learns about a Rubik
cube. She learns that it is a 3D puzzle, it has six faces and each has a solid colour
(but remember she has never seen colour before). There is an internal component
which allows the colours to turn and mix. The point of the puzzle is to return the Rubik cube to its original faces of colours. Mary learns all the general physical facts about a Rubik cube and finally upon her release, the guard at the door hands her a Rubik cube. Mary will then know particular physical facts about that particular Rubik cube she holds in her hand. Perhaps the corner of the cube is chipped or it does not turn with ease, which her general knowledge of rubric cubes claimed it would: universally quantified statements would state that Rubik cubes have an internal component which allows the colours to turn and mix with ease. She learns particular facts about the Rubik cube in her hand in a particular place and time as an indexical concept which she could not have learnt from what can expressed in the language of physics. It is possible that, with the proper description in the black and white room, she learnt particular facts of a Rubik cube with a chipped corner etc, but she would not have learnt that those facts pertained to this Rubik cube she holds in her hand.

4. Conclusion

In this chapter, I proposed a view I call bifactualism. Bifactualism is a physicalist account which suggests (amongst other things) that a physics account may be incomplete because there is no universally acceptable definition or account of what makes something physical and also because a good science is one that has expectations of failure and success. This means no good physicalist account
should claim to have knowledge of all things physical in this world and will readily accept that there are facts that may not be expressible in the language of physics.

There are two kinds of facts in the world: general and particular. A general fact is one which makes use of quantified statements which is expressible in the language of physics. Particular physical facts are not expressible in the language of physics and therefore escape the physicalist story. An example of a general physical fact is that all fish swim and a particular fact being that Nemo swims quickly.

Jackson claims that Mary learns all physical facts in the black and white room and I claim that she does not. It is wrong for Jackson to assert that from learning what is expressible in the language of physics, Mary is able to all the physical facts of the world. Physics cannot explain the particular facts about objects because it expresses its findings in generalities. It uses quantified propositions to express physical facts. Particular facts cannot be described using quantified statements. It may be the case that Mary needs to leave the room and come into contact with an object in order to be able to know particular physical facts. She will be able to say of a crippled cat: “this is what this particular cat looks like and it only has three legs when a quantified statement would claim that all cats have four legs”. A
particular physical fact however would claim that “this cat with three legs can be found at $p \text{ at } t$”.

Bifactualism was developed as a response to the Knowledge Argument but should not be considered exclusively as such. There is logical space for another account of experience through acknowledging that what can be expressed in the language of physics leaves out the account of particular physical facts. Particular physical facts are facts about consciousness and the bifactualist argues that this is the reason that physics cannot express facts about consciousness.
Chapter 4
The Bifactualist Response to the Mind-Body Problem

In this chapter I propose bifactualist responses to several issues that have been highlighted throughout this paper. In this chapter, I primarily focus on Nagel’s (1974) *What it is Like to be a Bat?* Nagel claims that we cannot know what it is like to be a bat because we can never have access to its subjective character of experience. This chapter argues that it is as difficult to know the feeling of what it is like to be a bat which has a feeling of what it is like: WIL (Prinz, 2012), as it is to know the feeling of what it is like to be a book (which has no feeling of what it is like: non-WIL). I argue that this is not because of two different ways of knowing two different properties, but rather that there are two different physical facts about both WIL and non-WIL properties. I show that with a bifactualist account, there are particular physical facts that can be known about WIL and non-WIL properties alike that are not expressible in the language of physics.
1. Three Contributions to the Mind-body Problem

In the first chapter of this paper, I discussed the role of neurobiological contributions to the mind-body problem. I explained that theorists in neurobiology hope to identify neural correlates of consciousness which will apparently give more insight into consciousness. The second notion is that the mind is more than a brain-process and it is possibly an entire embodied process. I explained that mapping the neural correlates of consciousness or claiming that consciousness is an embodied process still does not answer some pertinent questions in philosophy of mind.

In the second chapter I presented Jackson’s Knowledge Argument and discussed various pertinent contributions to the debate which included arguments from Lewis, Michael Tye, Coleman, Loar, Kuipers, and finally Churchland. Most theorists who challenge the Knowledge Argument either challenge the second premise or the validity of the inference but seldom do they examine the first, where I have sought to put pressure. In this chapter I briefly discuss questions stemming from the theories of NCC and the embodied process and put forward possible answers, using bifactualism. I also discuss possible questions stemming from the Knowledge Argument and put forward possible answers, again using bifactualism.
2. Neurobiology vs. Bifactualism

In the first chapter I discussed various neurobiological contributions to consciousness studies and in the second I focused on claims made about NCC and claims made about embodied processes. In particular I focused on contributions made by Damasio. I identified several issues with regards to the neurobiological undertaking and claims. In the following few paragraphs I am going to identify each issue and provide a bifactualist response or possible solution to each.

Among the first issues that came to the fore in the first chapter was the claim that there is a problem of interaction and causation. This has resulted in physicalists providing their own set of answers to the mind-body problem. This would entail a physicalist account which I have argued relies heavily on what is expressible in the language of physics. The bifactualist response to this physicalist account of interaction and causation is that there are certain facts about this kind of interaction that cannot be expressed in the language of physics. This however does not mean that because this interaction and causation cannot be expressed, it is therefore somehow non-physical or simply does not exist. It is possible that facts about the particularity of the interaction process as well as the particularity of the mental states in question are not expressible in physics.

A possible counterargument to this bifactualist response is that the bifactualist is making claims about facts but cannot tell us how this interaction takes place
either. The bifactualist account does not answer the problem of proving the physicality of mental states but assumes the physicality of mental states and also assumes the particularity of these mental states. The particular facts of mental states are not expressible in the language of physics. Therefore, the physicalist should not rely blindly on what is expressible in the language of physics.\(^\text{13}\)

The second issue I identified is that we have established that NCC will tell us that consciousness has neural correlates and that certain neuronal activity shows features or behaviour of consciousness. However, this brings us no closer in answering exactly how this consciousness is brought about or even the purpose of consciousness in the human body. It shows us that damage to certain parts of the brain impairs certain consciousness episodes. The bifactualist response to this issue is that it is possible that both general and particular physical facts exist about these episodes and given the history of physics, it is possible that physics will soon be in the position to grasp general facts (and perhaps some particular facts) about the generation of consciousness. However, the bifactualist will still claim that there are facts that will elude physical enquiry and will evade what can be

\(^{13}\)Again, it is important to note that the bifactualist does make some non-reductive physicalist claims but it does not make claims with regards to supervenience relations to counteract reductive physicalism. Rather, the bifactualist is making claims regarding what can be expressed in the language of physics and is not making claims regarding supervenience relations. The bifactualist claims that all facts are physical: some are general and some are particular.
expressed. Therefore, we will always have an incomplete knowledge of these things.

A possible counterargument to this response is to claim that the bifactualist is no closer to explaining how consciousness is generated than either the dualist account or the neurobiological account. The bifactualist could respond to this counterargument by again claiming its intentions is not to prove the existence of things but rather to say that there is a bifactual account to be had.

The third issue I identify is that Damasio and others proponents of embodied cognition have argued for feelings and the role of feelings in consciousness. However, they cannot show us how these feelings are generated, and how they are demonstrated or presented. The bifactualist response to this would be the same as the response above advocating that physics may have the answers to these things in the future, however there will still be some facts that escape the physics account. Also, feelings are particular to an object and we have already identified that physics does not have a grasp of this kind of particularity.

The bifactualist account, at first glance, seems to be similar to claims made by Colin McGinn as they both argue that certain facts cannot be known to certain persons or certain forms of enquiry. McGinn states that even though “we know that brains are the de facto causal basis of consciousness, [...] we have, it seems, no understanding whatever of how this can be so” (McGinn, 1989: 395). His
explanation for this is because of cognitive closure: “A type of mind $M$ is
cognitively closed with respect to a property $P$ (or theory $T$) if and only if the
concept-forming procedures at $M$’s disposal cannot extend to a grasp of $P$ (or an
understanding of $T$)” (McGinn, 1989: 385).

McGinn further claims that “cognitive closure with respect to $P$ does not imply
irrealism about $P$. That $P$ is (as we might say) noumenal for $M$ does not show that
$P$ does not occur in some naturalistic scientific theory $T$ – it shows only that $T$ is
not cognitively accessible to $M$” (McGinn, 1989: 296). This means that
consciousness exists as a physical property and the question should not focus on
establishing the physical nature of consciousness, but rather how we begin to
access or establish any theory of consciousness given that we are cognitively
closed from it. It is our own mind’s inability to grasp a certain property $P$ which
does not allow us to know or establish a theory $T$ about consciousness, even
though that property $P$ is real and in no way non-physical. The truth for McGinn
is that “a problem that strikes us as deeply intractable, as utterly baffling, may
arise from an area of cognitive closure in our ways of representing the world”
(McGinn, 1989: 396).

The bifactualist account agrees with McGinn’s claim that we are unable to grasp a
certain property but it shies away from claiming that “$T$ is not cognitively
accessible to $M$” but rather that it is the way of representing facts about
consciousness that proves to be tricky. Of course, the bifactualist in no way claims that the individual herself will know all the particular physical facts about consciousness occurring in her own body, but she will most certainly be able to grasp some facts about her particular consciousness that physics just cannot quite express. The bifactualism rejects McGinn’s overall theory of Cognitive Closure because McGinn claims that the mind is not able to grasp certain facts. The bifactualism, in contrast, is suggesting that the mind is not cognitively closed off, but that physics is.

Dennett (1991) argues that McGinn is quick to suggest a cognitive closure but fails to explain how we are able to ask the questions about the concepts in the first place. An example of a monkey anxiously wanting to know more about electrons is ludicrous because monkeys truly are cognitively closed off to the idea of electrons. Then it would seem that we are not closed off to the concept of a mind, since we are able to formulate questions about the mind. Dennett (1991: 3) writes that “we should be unimpressed by the example of the monkey, to whom the electron is out of bounds, for not only can it not understand the answers; it can't understand the questions. The monkey isn't baffled, not even a little bit”.

For bifactualism, a kind of closure is a problem regarding facts about the mind. However, the closure does not arise from our being cognitively closed off to our own minds, but rather that physics is unable to express certain facts about the
mind. Because we are able to formulate questions about the mind, we are not
cognitively closed off to the mind. Physics cannot express particular physical
facts about the mind and therefore it is possible that physics is closed off.

3. The Knowledge Argument vs. Bifactualism

In the Knowledge Argument, Jackson makes dualist claims such that Mary learns
all physical facts in the black and white room and then Mary comes to learn non-
physical facts once she escapes the room. I have already identified that premise 1
of the argument is problematic. Most theorists challenge the second or third
premise and I challenge the first. In this section I am going to very briefly identify
issues that spring from responses (Lewis, Tye, Coleman) to the Knowledge
Argument and provide a possible answer through bifactualism.

In response to the Knowledge Argument, Lewis claims that Mary learns abilities
and no new facts. The bifactualist response to this is that these abilities would still
have facts about them and these facts are particular physical fact regarding
remembering, recognising and imagining. Physics cannot communicate these
particular physical facts and they are certainly not expressible in the language of
physics. This is why Mary could not know them in the black and white room. If
there are facts to be known about remembering, imagining and recognising
objects, then Mary cannot know these facts in the black and white room.
Tye argues that Mary only knows certain things as indexical concepts and is not able to remember very specific objects or concepts such as red17 and therefore does not learn any new ability or fact. I find this problematic as it seems improbable that the inability to remember facts means the facts are somehow not learnt or do not exist. The fact that red looks a certain way exists regardless of whether or not a person is able to remember it. There are facts about indexical concepts and there are facts about red17.

It is possible that Mary learnt this new particular fact of seeing red17 that she could not have known in the black and white room, but she fails to remember it. Therefore, Mary does not learn this particular physical fact in the black and white room but learns it upon her release. The existence of the fact should not be questioned because she is unable to remember it. An example is if I ask you to tell me exactly what colour your coffee was this morning. You probably will not be able to tell me. But when you drank it this morning, there were facts about your coffee that you knew then (at a particular time and place) and forgetting the colour now does not mean that facts about the colour of coffee do not exist.

Churchland, Kuipers and Coleman argue for knowledge-by-acquaintance and this is a very interesting point. However, being acquainted to something entails facts about that acquaintance. The same response to Tye applies here as well. The bifactualist would argue that facts about knowledge-by-acquaintance are
particular and are physical. A possible response to the bifactualist would be the fact that it is more difficult to communicate fact about knowledge-by-acquaintance in comparison to a layout of a person’s particular office. The bifactualism would answer that both need to be understood as indexical, regardless of whether you are comparing books in an office next door, I would still need to situate myself in a time and space to be able to understand it as a referential term. Experience is not necessary to know particular physical facts, but knowing these facts as indexical is.

The crucial issue which I would like to devote a section to is the problem of access. The problem of access concerns how we are able to access mental states and subjective character of experience. I would like to focus my attention on the subjective character of experience and tentatively suggest that bifactualism could explain certain aspects of subjective experience and access. Please note that this is only a tentative suggestion and I do not claim that bifactualism has the answers. Nonetheless, I do think it might have some interesting contributions which I put forward here.

4. Nagel on the Problem of Access

Bifactualism attentively wants to answer the problem of particular access and I here discuss Nagel’s hypothesis in more detail in order to discuss the philosophical problem of access. Nagel’s (1974) article: *What it is like to Be a Bat*
discusses the concept of subjectivity and shows how subjectivity is not transferable nor can it be understood through third-person enquiry. Nagel discusses the idea of consciousness and experience by suggesting that we could never know what it is like to be a bat because we are not conscious of the bat’s experience nor do we share the subjective character of its experience. The problem Nagel draws attention to is that every subjective character is unique and is part of a subjective phenomenon which is specifically of a single point of view and is accessible only by the person or thing experiencing it. Chalmers (2000: 197) writes that “Nagel focuses attention on this aspect of consciousness and argues that it is particularly difficult to explain [...] we might know all about the objective functioning in a bat’s brain, but we still would not know what it is like to be the bat, from its own subjective viewpoint”.

Nagel imagines that the bat undergoes mental events or states. The argument entails attempting to imagine what it would be like to be a bat, and according to Nagel this is impossible. Nagel assumes the bat has a feeling of what it is like or a subjective character of experience. Even if we were to imagine flying around at sunset snacking on insects and relying on a complex sonar system as a form of perception, we would only succeed in imagining what it is like for us to be a bat, but we would be no closer to knowing what it is like for a bat to be a bat. If we try to establish some kind of objective account of its subjective experience, we
move further away from knowing the experiential quality of the bat’s embodied
cognitive events. “If the subjective character of experience is fully
comprehensible only from one point of view, then any shift to greater objectivity-
that is, less attachment to a specific viewpoint—does not take us nearer to the
real nature of the phenomenon: it takes us farther away from it” (Nagel, 1974:
223).

This is the crux of the problem: how do we go about accessing the feeling of what
it is like or ‘WIL’ (Prinz, 2012) to be a bat? The understanding is that if we
cannot access the WIL of a bat, then somehow our knowledge of a bat is
incomplete. It seems that if we cannot access the feeling of WIL then we remain
ignorant of the bat’s WIL. It may actually not be good enough that we can know
facts about WIL. Knowing facts is not the same as feeling facts it seems. It is
almost as if we insist that in order to know a bat completely, we would need to
feel like a bat and it can be argued that knowing and feeling are two different
exercises and should not be used interchangeable just because we introduce WIL
and non-WIL into the equation. So there are two problems here.

First, it seems that access through knowing WIL facts about something which has
mental states may not be good enough to fully understand the mental states of that
organism. Second, it seems as if some hold that access through feeling WIL facts
is a better way to understand something which has mental states. I am going to
call this the traditional view of access. The traditional view of access claims that bats have mental states and learning all the available facts about those mental states leaves something out of the story and the only way we can truly know those mental states is to feel those mental states in the same way a bat does. I want to challenge the traditional view of access by claiming that learning facts about things which have WIL should be a similar exercise to learning facts about things which have non-WIL.

5. Access to WIL and Non-WIL

According to Nagel, in order to understand the subjective character of experience, we need to know the feeling of WIL to be a bat. Bats aside, I am going to ask you to imagine what it is like to be something without any mental states and thus no subjective character of experience at all. I will call these objects non-WIL. Let us say I ask you to imagine being something as inanimate and non-conscious as a book. You cannot. The reason is not because we cannot know its experiential quality. It does not have any experiential quality for us to know. Nor is the problem that we do not have enough information about it. All the general and particular facts about the book are available to us (we may suppose). The real reason we cannot know what it is like to be a book is because we are going about this access problem in completely the wrong way. Nagel argues that “if
physicalism is to be defended, the phenomenological features must themselves be
given a physical account” (1974: 220).

I am led to ask what type of physical account would be sufficient for us to believe
we have enough knowledge of something that has subjective experience? We feel
that having knowledge of a book is good enough (because it has no experiential
quality), yet we are not content with using the same kind of physicalist approach
to label or define subjective experience in fear that we will not capture an
important aspect of the subjective experience of that thing. There seems to be an
idea that in order to properly understand an organism, we must be able to feel
what the other feels like. And because we cannot feel what another feels, we can
therefore never really know and thus our knowledge is somehow incomplete.

This criterion for knowing WIL is not a stipulated condition for knowing
properties which are non-WIL. Yet we feel that we properly understand what a
book is. I argue that access to the possible mindfulness of a book and the
mindfulness of a bat is not the problem here. The problem is the different kinds of
knowing we imagine we have. There appears to be two different kinds of
understanding: through knowing and through feeling. To differentiate two
different kinds of knowing is the problem here and this idea that there are two
distinct ways of knowing things has led to the development of the traditional view
of access. Knowing facts about WIL and facts about non-WIL, I tentatively argue, may in theory, be about the same kinds of facts.

A possible counterargument to the above claim is that WIL facts and non-WIL facts are not the same. Bats have WIL and we cannot access facts about the bat’s WIL. A book has no WIL, so there are no facts about a book’s WIL to access. This shows that WIL facts are different from non-WIL facts with regards to the way we are able to access them. And here lies the problem of access to particularity of facts. The next section is devoted to how the bifactualist could respond to this problem.

6. The Bifactualist Response to the Problem of WIL and non-WIL

There are general physical facts that can be known about WIL if we can concede that WIL are neurobiological processes. There are also general physical facts about non-WIL and then there are particular physical facts that can be known both about WIL and about non-WIL. I am choosing to describe conscious mental states (WIL) in the same sort of manner as other physical properties (non-WIL) even though I do concede that it is possible that they may not be exactly the same type, I am claiming that they are both physical and therefore the approach should be similar. There are facts about each that can be known by both physics and by a person encountering the particularity of both WIL and non-WIL. Physics can describe WIL and non-WIL by providing general physical facts about them.
Physics can tell me about NCC and embodied cognition. Physics cannot however provide particular physical facts about them. When bifactualism is applied, we do not question the physical-status of WIL or non-WIL properties, but rather we understand that there are facts about particularity that are missing in both the WIL and non-WIL physicalist account.

Facts that can be expressed in the language of physics are not facts about particularities. There are particular facts about WIL and non-WIL that physics cannot tell us about. Physics can tell me about books in general and physics can tell me about mental states in general (through neurobiology). Physics cannot tell me anything particular about books, it cannot tell me that you have books on your bookshelf. Physics can also not tell me anything particular about mental states. It cannot tell me about the particular neuronal activity occurring in your brain as you read this. It can offer a description that fits that activity but it cannot assert that that is what is in fact going on.

I argue that the enquiry for both WIL and non-WIL could be similar. There are particular physical facts about books as there are particular physical facts about conscious mental states. Again, I am not quite comfortable with saying that the particularity of conscious mental states can be known in exactly the same way as knowing particularity of a book, but in theory, the enquiry or knowing thereof is similar in that we learn particular physical facts. Now the actual problem of
access comes into play. The traditional view of access was that we needed to feel what another felt in order to know it better. I argue that learning facts about WIL and non-WIL is probably a similar process. The way we learn the particularity of a non-WIL property is possibly the same way we learn the particularity of our own WIL.

Particular physical facts are known through proper names and indexicals. Thus I claim there has to be some kind of spatiotemporal relation to the object known. In order to know particular physical facts about the book on your table, I would need to know that you have an office and that this particular book can be found there now. If you tell me you have a copy of *Wuthering Heights* on your table, I am assuming the book is there at this moment. The particular fact of your having *Wuthering Heights* on your table is dependent on time and space and my indexical claim I am able to make. Because you told me of the book, I am able to make the indexical claim that “there is a book on your bookshelf”. Conscious mental states are a bit more troublesome given that to know a book, I can have an indirect relation to it but with mental states, I must have a direct relation to the states itself in order to know its particularity.

The big question we should ask is not how we are able to have access to WIL to learn facts about WIL, but how we learn particular physical facts of both WIL and non-WIL since I argue the issue of knowing particularity is similar on both
accounts. In the following section I tentatively offer an account of how it is possible for a person to know particular physical facts about her own WIL and other non-WIL.

7. What Mary Actually Learns

We have looked at what constitutes general physical facts and Mary, in the black and white room learns these types of facts when learning about the colour red and its effects on human vision. Remember that these general physical facts are facts about physical properties and laws. This kind of fact is generated from events, states and objects and it is these facts that Mary acquaints herself with in the black and white room. Mary spends her time in the black and white room learning about what other people experience when seeing the colour red but what exactly (according to bifactualism) does Mary learn about somebody else seeing red?

By learning about the colour red and the effect it has on the human eye and how organisms process and interpret the sense-data they receive, Mary learns general physical facts about these physical properties. The kind of learning that Mary undergoes in the room involves general physical facts and in order for her to acquire these general physical facts, she would have needed to do the following (and possibly more):
Mary would have needed to 1) view MRI scans to chart or map the colour pathways of the normal brain beyond V4\(^{14}\). The general physical facts Mary acquainted herself with during this exercise was that the colour pathways is a process at a specific brain site. Mary could only do this once 2) she had monitored how brains would react to colours as properties. She would need to know 3) how the wavelength of individual colours reflects off surrounding surfaces and again how brains are able to decipher this information. Mary would learn that the process of seeing colour does not only just take place in the V4 region of brains but really involves two other regions too: the V1 and V2 regions of brains (Zeki and Marini, 1998: 1669). Mary learns the following (and more):

a) Brains have pathways
b) Colours as properties have a causal relationship on brains
c) The causal relationship can be seen by the way the colour pathways become active in brains when colour is seen reflecting off surrounding surfaces.

We will call this process that Mary just examined: \(abc\). This is a physical occurrence explainable using general physical facts of people seeing red. We can see that Mary has learnt many physical facts (of course a lot more than listed above), but there are a few facts she has not learnt. She does not learn the

\(^{14}\) There are different layers of primary visual cortex labeled V1 to V4 (Prinz, 2012: 296)
particular physical facts (which are not expressible in the language of physics) of these physical properties occurring as an observer of a particular spatiotemporal event \( x \). Mary cannot know these particular facts unless she herself occupies the correct relation to the object. As mentioned before, knowing particular facts entails occupying a certain relation to the object in order to know the indexical.

8. **Red: According to John**

In the previous section, Mary examined brains, colour perception and gathered some general physical facts of what she thinks someone else’s experience is about. Let us assume one of those brains belonged to a man called John. Mary has learnt all these general physical facts about John’s brain. What she does not know is John’s particular experience as an observer in an event \( x \{p1+t1\} \). Let us tell John’s story from the observer of event \( x \) and see what kind of particular physical facts Mary could learn if she herself occupies the role of the observer of the event \( x \{p1+t1\} \). For whatever reason, John finds himself in a poppy field. John is struck by the beauty of all the red poppies, especially in contrast with the bright green leaves. John may not be overtly aware of his bodily processes at that particular moment, but we should be aware that he is experiencing seeing the colour red from a constrained bodily condition (Embodied Cognitive Perspective). John is seeing red and according to Mary’s scientific expertise and working with the general physical facts she is privy to, John is experiencing \( abc \).
However, John is also experiencing the seeing of red from a constrained bodily-state and doing so in a very particular way. He is experiencing the poppy fields as a particular ECP property of the event $x$ at a spatiotemporal particular. Facts about this event are also particular. These particular facts pertaining to the object can be learnt by having a relationship to the object perceived or encountered. John can have access to these particular physical facts because he occupies the role of the ECP in the event $x \{p1+t1\}$ encountering the object (also a property of the event). This is a possible condition for knowing particular physical facts of WIL. This would mean that Mary, given all her knowledge of general physical facts cannot know the particular physical facts of John seeing red because firstly, she is not occupying the ECP of the event $x \{p1+t1\}$, nor does she have a relationship to the particularity of the object. Thus, Mary cannot know the particular physical facts of John’s ECP event $x \{p1+t1\}$.

A possible objection to the above claim is to state that we are able to know particular physical facts about a book but we still remain ignorant as to the particular physical facts of John seeing red. As mentioned before, we need to have a relation to the particularity of an object. We need to somehow have a reference point to books in the vicinity at a particular time and place for us to know them as particulars. I must be able to say: “next door there is a bookshelf and on this bookshelf is a book called Wuthering Heights”. These books must occupy a
spatiotemporal sphere for us to learn facts about them. Now, WIL are neurobiological processes and unless we have some kind of reference point to these neurobiological processes in the particular spatiotemporal sphere, we will not be able to express any particular fact about WIL and can therefore not know the WIL of another, similar to how I cannot know particular physical facts about the books on your bookshelf because I have no reference point to do so. I have no reference point to your WIL, therefore I cannot know its particularity the way you do.

Particular physical facts learnt through events can be known by the observer. I argue that Mary, when she leaves the black and white room and experiences the colour red for the first time, does so as an observer in an event. Lewis claims that no new knowledge is learnt but only an acquisition of abilities. These include the ability to recognise, imagine or remember and if these take place as a particular physical event, then there are particular physical facts generated from this experience which can indeed be learnt. Mary learns the properties of event $x$, regardless of whether she learns abilities or the feeling of WIL to see the colour red. Lewis states that “knowing what it is like isn’t the possession of information at all” (Lewis, 1983a: 131). I disagree with this statement. Knowing WIL is learning facts about a particular role as an observer.
If bifactualism is right, then Mary, when learning about herself as a property of an event, learns a new kind of physical fact. If she learns the abilities to imagine, remember and recognise, she is doing so from an ECP property within an event and therefore certain particular physical facts are available about the event. Lewis is claiming that Mary learns something new upon her leaving her room, but it is not a new kind of knowledge. I claim that if Mary is learning anything about herself as a particular property of an event, then we can say she is learning a new fact and this fact is a particular kind of physical fact about herself and about the object observed and therefore she does learn new physical facts.

9. Conclusion

The Knowledge Argument is a contentious debate for philosophers of mind. I believe that this thought experiment is especially arguable because of the way it was set up and have argued this extensively in the last two chapters. The first premise of the argument states that Mary learns all physical facts in the black and white room and through bifactualism, I have shown that this is impossible. In this chapter, I defend bifactualism by responding to several issues I have identified throughout this paper. I give a possible bifactualist response to issues arising from both NCC and embodied cognition. I then give a bifactualist response to Lewis, Tye and several theorists advocating knowledge-by-acquaintance.
Nagel argues for a subjective point of view in both his *What is it like to be a Bat* and *A View from Nowhere* (1986). This subjective character of experience is what some philosophers feel will never be accessed or known in the same way that we know facts of objects or things which have no subjective character of experience. I argue that facts about *what it is like* (WIL) are about particular physical facts and should be addressed similarly to particular physical facts about objects which have no *feeling of what it is like* (non-WIL). I argue that both WIL and non-WIL are physical and have general and particular physical facts. Physics is able to tell us general physical facts about both WIL and non-WIL properties and it is unable to tell us particular physical facts about both WIL and non-WIL properties.

We seem to think that we can understand properties (without any subjective experience) from a purely objective, scientific account, but we do not want to use the same kind of approach for something that has subjective experience (WIL) in fear that we lose some kind of fundamental element of it. I argue that if this is the condition by which we can fully know something, then we lose a fundamental element of even the most non-WIL properties because we cannot know what it is like to be something which has a subjective character of experience, let alone know what it is like to be something which has no subjective character of experience at all because these facts are particular physical facts and is not ours to know through what is expressible in the language of physics.
I conclude that a person occupies a role in a spatiotemporal event and that in order to have knowledge of certain particular physical facts of this event, a person would need to have a relationship with the object. I argue that no two persons can occupy the same properties of the same spatiotemporal event and therefore Mary, in her black and white room can never know what it is like to be John because she does not know the particular physical facts John can know in relation to the object. Mary, in her room learns all general physical facts of the colour red and when she leaves the room and sees red for the first time, she then learns the particular physical facts of seeing the colour red. I consider Mary’s knowledge of all physical facts as incomplete in the black and white room. Thus I claim that Jackson’s hypothesis is flawed: Mary does not learn all physical facts in the black and white room. We need to have a bifactual account of experience and can only do this by learning general physical facts and particular physical facts.
Conclusion

Jackson challenges the physicalist by suggesting a highly-debateable thought-experiment called the Knowledge Argument. The Knowledge Argument is so highly-debateable because the physicalist struggles to answer the question of what facts, according to Jackson, escape the physicalist story. The physicalist claims that if a fact escapes the physics story, then it does not exist. Jackson argues that if a fact” escapes a physics story, then it is non-physical. I argue otherwise. I argue that the physicalist needs to review its dependence on what is expressible in the language of physics. There are facts which are not included in the physicalist story, not because they are non-physical or because they do not exist. These facts are not included in the physics story because there are certain facts which are not expressible in the language of physics. The physicalist needs to expand its horizons and by extension, review its dependence of what can expressed in the language of physics. The way to do this is to admit that physics cannot express all physical facts available.

In this paper, I argued for a physicalist account called bifactualism. Bifactualism claims that there are two kinds of physical facts. The one kind of fact is what physics is capable of expressing (the general physical fact) and the other kind of fact (particular physical fact), to which physics is closed off to. I do not make any
claims to have fully answered the pressing questions in the Knowledge Argument nor do I deny the important of neuroscience in this debate. I hope, through this paper, to have contributed an idea to philosophy of mind and especially to physicalism and perhaps provoke physicalists to work on a model to strengthen itself as well as open itself to a world wider than what physics can currently represent.

Traditionally the physicalist holds that all things that can be expressed in the language of physics are physical and anything else which cannot be expressed in the language of physics, does not exist. With this kind of statement, I argue that the physicalist has limited its claim on what is physical in this world. In this paper I argue that there are certain physical facts which are not expressible in the language of physics and these facts should be included in an account of physical facts. With the ever-changing progression and, at times failure, of physics enquiry and discovery, I suggest that the physicalist review its stand that all things describable in the language of physics exist and all other things are false or do not exist.

A physics which actively contributes to this physicalist account is neuroscience. Neuroscience has elbowed its way into several philosophy of mind debates and hopes to prove (within a physics account) that such things as consciousness, phenomenal states, subjective character of experience etc., is either reducible to a
physics account or at least emerges from physics. As discussed in the first chapter of this paper, such claims are problematic. Neuroscience is struggling to prove or disprove physicalist claims, and struggles further to completely reject dualism. Neuroscience, at times, fails to provide evidence which will be regarded as satisfactory for a physicalist account. The physicalist further chooses to discard what is not expressible in the language of physics. This limits any kind of dualist modes of explanation; however it cannot satisfactorily deny claims made by the dualist. Bifactualism is a physicalist account which, in some roundabout way, is compatible with some dualist claims. Bifactualism also allows physicalists a stronger claim on the physicality of things in the world by not limiting their dependence on what is expressible in the language of physics.
Reference List


